

Madeleine ARNOLD

en collaboration avec **Claude LEBRUN**

Lutèce'IA, Paris

REPRÉSENTATIONS LINGUISTIQUE et SPATIALE d'une SCENE TRIDIMENSIONNELLE

Remarques à propos d'une expérimentation en synthèse d'image

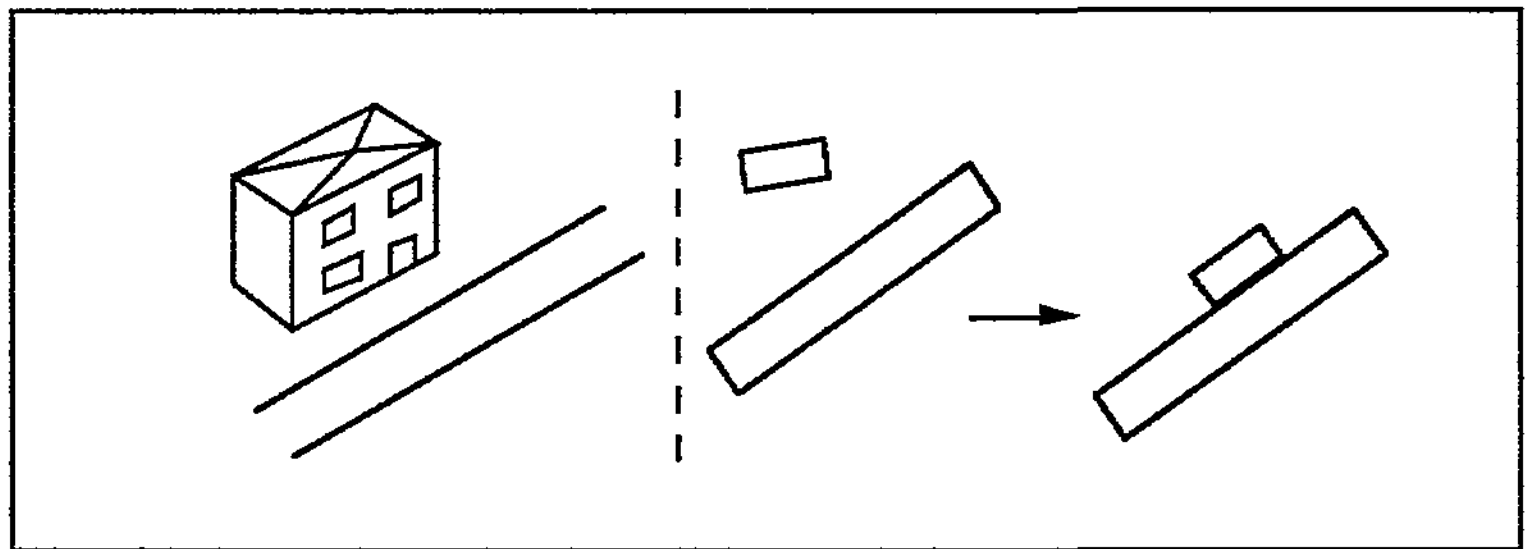
Introduction

L'invention de l'espace dans un environnement de conception informatisé prend différentes formes. Le système informatique obéit pas à pas aux instructions élémentaires que lui donne le concepteur et se contente d'effectuer des calculs ou de construire des représentations graphiques à partir d'éléments primitifs —points, segments, facettes, volumes. Il peut aussi remplacer le concepteur pour certaines tâches routinières telles que coter un dessin ou élaborer une maquette tridimensionnelle à partir d'un plan et de quelques informations supplémentaires sur des hauteurs en particulier. Pour d'autres tâches moins routinières, on attend du système qu'il fournisse une assistance intelligente, qu'il soit capable, par exemple, de retrouver dans sa mémoire des schémas de solutions antérieures et de les appliquer à une situation nouvelle [Guéna, 1992]. L'un des aspects d'une assistance intelligente au concepteur, qui commence à retenir l'attention dans le domaine de la recherche en CAO (conception assistée par ordinateur), est la capacité de dialoguer d'une façon conviviale. La *multimodalité* apparaît alors comme un caractère important des interfaces homme-machine. L'utilisateur, ou des utilisateurs travaillant en collaboration sur un projet donné, doivent pouvoir dialoguer avec le système en employant différents modes d'expression —une langue écrite ou orale, l'image, le geste— isolément ou conjointement.

Une sorte d'interface multimodale est celle qui permet de transcrire un texte en des images. Nous nous sommes demandé si une telle interface ne contribuerait pas à faciliter la communication entre les concepteurs et les

¹Le menu se présente en général comme un tableau ou une bande compartimentée dont chaque case contient une icône, une expression verbale ou un échantillon, indiquant un choix possible parmi des formes ou des opérations géométriques, des couleurs, des lumières, des textures, etc. L'utilisateur indique, à l'aide d'un curseur, par exemple, quelle case il choisit.

systèmes de CAO ou de synthèse d'image. En effet, les interfaces actuelles sont peu sémantisées. Ainsi pour créer une scène architecturale ou urbaine tridimensionnelle, un utilisateur est obligé de passer par des menus¹ combinés à des instructions tapées au clavier, ou bien plus rarement par un langage formalisé. Par l'entremise des commandes dont il dispose, il manipule des cubes, des parallélépipèdes, des cônes, etc. et effectue des translations, des rotations, des homothéties, etc. Les objets qu'il traite ne sont donc pas des maisons, des fenêtres, des escaliers, des murs, ou autres entités fortement sémantisées, mais des formes géométriques complexes composées de formes géométriques de base. Il ne peut pas demander que les maisons soient mises au bord de la rue ou que la fenêtre soit agrandie ou déplacée vers la gauche, par exemple. Il n'est pas possible pour le concepteur de travailler directement avec des entités et des opérations qui lui sont familières. Il doit sans cesse traduire ce qu'il veut faire dans des commandes compréhensibles par la machine. Cette situation n'est pas obligatoirement un handicap pour un concepteur expérimenté. Avec l'expérience, des automatismes se créent et le concepteur sait, pratiquement sans réfléchir, quelle séquence de commandes et quelles valeurs donner à des paramètres pour obtenir un certain effet. Par ailleurs le fait de travailler sur les primitives qui composent les objets et les opérations peut être un facteur de plus grande liberté de création. Il reste que la pauvreté sémantique des commandes dont dispose actuellement le concepteur introduit des lourdeurs de fonctionnement manifestes. De plus elle rend difficile sinon impossible certaines améliorations *intelligentes* en matière d'organisation des connaissances et de raisonnement. Prenons, par exemple, le cas d'un système qui doit mettre une maison au bord d'une rue et qui dispose d'une procédure qui lui permet de mettre un rectangle contre un autre rectangle, situé dans le même plan.



Actuellement un système est incapable de distinguer sémantiquement les rectangles qui constituent le fond de ceux qui correspondent aux murs de la maison ; de même le ruban rectangulaire qui représente la rue n'est qu'un rectangle parmi d'autres. L'utilisateur doit donc indiquer quel rectangle doit être mis contre quel autre rectangle. Dans un système à la

sémantique plus riche qui saurait distinguer une maison d'une rue et le fond des murs de la maison, et qui saurait également que pour placer un objet tridimensionnel de type *bâtiment* au bord d'un objet bidimensionnel de type *voierie*, il faut mettre le fond du bâtiment contre l'objet de type *voierie*, il suffirait de demander que la maison soit mise au bord de la rue.

L'utilisation d'une interface texte-image dans un environnement de conception pose un certain nombre de problèmes *a priori*. En effet, quatre types de systèmes de représentation et de traitement des connaissances sont impliqués dans la construction d'une scène : une langue, l'image mentale, un langage graphique, et un langage de commandes. L'image mentale joue un rôle en amont et en aval de l'emploi de la langue. En effet, l'utilisateur qui décrit un fragment de scène ou qui donne un ordre de modification de celle-ci vise un effet spatio-visuel qu'il anticipe mentalement à travers des images plus ou moins nettes et élaborées. Une fois l'énoncé verbal produit, il revient à la machine de le traiter. Mais pour qu'elle puisse le faire, il faut que l'utilisateur ou une autre personne ait implanté les mécanismes d'analyse linguistique et de création infographique² nécessaires.

²'Infographie' :
production de
graphismes ou
d'images à l'aide de
l'informatique.

Le concepteur de ces mécanismes, que nous appellerons "l'ingénieur" pour le distinguer du concepteur de la scène, effectue la démarche inverse de celui-ci. Le concepteur part des images mentales qui occupent son esprit et leur donne éventuellement une formulation linguistique. Par contre, c'est l'énoncé verbal qui est donné à l'ingénieur auquel il revient d'analyser et d'interpréter infographiquement les représentations mentales imagées attachées aux expressions linguistiques. Cette partie de son activité est essentiellement un travail de modélisation préalable à l'élaboration des mécanismes informatiques. La modélisation comprend aussi une analyse linguistique de l'énoncé verbal au cours de laquelle sont extraites de celui-ci les informations utiles à la construction de la scène. Ce n'est qu'une fois les modèles linguistiques et graphiques définis, ainsi que leur articulation, que l'ingénieur passe à l'élaboration effective des mécanismes —étant bien entendu que le savoir sur les types de formalisation informatique disponibles a une influence sur les diverses modélisations. La modélisation s'effectue par ajustements entre ce qui est imposé par les objets à modéliser —les représentations mentales imagées et l'énoncé verbal— et les possibilités de représentation et de traitement du système. La conception de mécanismes de transcription automatique d'énoncés verbaux en images souligne une question fondamentale, de nature sémiotique. Il s'agit de la rupture radicale entre les deux types d'ensembles sémiotiques que sont les langues et les images : l'expression verbale "une pipe" n'est pas le dessin de la pipe, pour plagier l'œuvre célèbre de Magritte. Pour passer automatiquement du modèle d'un énoncé

verbal au modèle de la représentation graphique correspondante, il faut établir un lien arbitraire explicite entre les descriptions informatiques des deux modèles.

³Texte de référence :

*phr. 1• L'église est
au centre de la place.*

*phr. 2• La mairie
est à gauche de l'église.*

*phr. 3• La place est
entourée d'arbres.*

*phr. 4• La rue
arrive à la place.*

*phr. 5• Des
maisons bordent le côté
droit de la rue.*

*phr. 6• A gauche il
y a trois immeubles.*

*phr. 7• Il y a un
magasin entre deux
immeubles.*

*phr. 8• Une voiture
est sur la place.*

*phr. 9• Une autre
voiture est au centre de
la place.*

*phr. 10• Trois
voitures roulent dans la
rue.*

*phr. 11• Un piéton
marche sur le trottoir de
droite.*

Une expérience

Pour préciser les problèmes de modélisation que pose une interface texte-image, nous avons procédé à une expérience. Nous nous sommes placés dans la situation d'un concepteur qui voudrait obtenir un schéma de scène tridimensionnel rapidement, en décrivant la scène verbalement. Le texte que nous nous sommes donné compte onze phrases décrivant une scène urbaine³. Dans chaque phrase, un objet est positionné par rapport à un autre. Le texte, en entier ou phrase après phrase au choix, est entré au clavier. Chaque phrase est analysée et le fragment de schéma de scène correspondant produit. Le processus de transcription est en partie automatique et en partie interactif. Le système de synthèse que nous avons employé est IKOGRAPH, réalisé par Michel Bret au CIMA [voir Annexe 1, p. 28]. Il permet la création, la modification et la visualisation de scènes tridimensionnelles à l'aide d'un langage évolué composé de formules alphanumériques⁴. Une scène est une maquette virtuelle tridimensionnelle [Clayssen *et al.*, 1987] dont la description formalisée est enregistrée dans la mémoire du système. L'utilisateur n'a pas accès à cette maquette ; il ne peut en voir que des images affichées à l'écran. Les objets composant une scène sont représentés par des surfaces polyédriques, dites couramment "à facettes". Ils sont des "formes" parmi d'autres, telles que les lumières, les points de vue, les brouillards ou les trajectoires. Les formes sont créées, supprimées, modifiées, visualisées ou déplacées à l'aide des commandes graphiques. D'autres commandes permettent, entre autres, de visualiser une scène selon un point de vue et un type de perspective déterminés, de traiter des chaînes de caractères et d'obtenir des informations sur des éléments de la scène. Il est possible aussi de rédiger des procédures — autrement dit des suites de commandes — qui, une fois appelées et exécutées, créent des formes ou agissent sur celles-ci.

⁴'Formules
alphanumériques' :
formules composées de
caractères
alphabétiques et
numériques.

Transcription du texte en une maquette virtuelle à l'aide d'IKOGRAPH

La transcription compte deux volets principaux — l'analyse linguistique et la synthèse infographique — qui sont articulés par l'entremise d'un *moule* et de *dictionnaires de correspondances*. Dans le moule — une formule canonique constituée de variables —, sont versées les informations nécessaires à la synthèse infographique.

Composition du moule :

*Entité1 Quantité1 Qualificatif1 Relation Entité2 Quantité2
Qualificatif2 Entité3 Quantité3 Qualificatif3.*

Relation est la relation de positionnement. *Entité1* est l'objet à positionner, *Entité2* et *Entité3* les objets de référence par rapport auxquels le premier objet est placé. Pour toutes les phrases —sauf une où la relation est "être entre"— il y a un seul objet de référence, qui correspond à *Entité2*. *Quantité1* a pour valeur le nombre d'objets à positionner, *Quantité2* et *Quantité3* les nombres d'objets de référence. Aux variables *Qualificatif* sont attribuées des informations complémentaires extraites par l'analyse linguistique ; c'est le cas de "autre" dans la *phr. 9* • *Une autre voiture est au centre de la place.*

Les dictionnaires de correspondances sont de simples variables dans lesquelles les informations sont enregistrées à la file. On en compte deux : *DictEnt* pour les entités et *DictRel* pour les relations. Dans *DictEnt*, à chaque identificateur linguistique d'une classe d'objets sont associés les identificateurs infographiques des occurrences correspondantes. Le principe est le même pour *DictRel* où l'identificateur linguistique d'une relation est lié aux identificateurs infographiques de la procédure permettant de construire cette relation.

Exemple :

Lorsque le système est parvenu à la fin du traitement des quatre premières phrases :

— *L'église est au centre de la place. La mairie est à gauche de l'église. La place est entourée d'arbres. La rue arrive à la place*—,

les contenus des deux dictionnaires sont les suivants :

• *DictEnt* = *église églisel / place placel / mairie mairiel / arbre arbre1 arbre2 arbre3 arbre4 arbre5 arbre6 arbre7 / rue ruel* —

— *Eglise, place, etc.*, sont les identificateurs linguistiques de classes d'objets. *Eglisel, placel, etc.*, sont les identificateurs infographiques des formes dénommées église, place, etc.

• *DictRel* = *etre-au-centre-de macetraucentrede / etre-a-gauche-de macetragauchede / entourer macentourer / arriver-a macarriver*

— *Etre-au-centre-de, etre-a-gauche-de, etc.* sont les identificateurs linguistiques de relations et *macetraucentrede, macetragauchede, etc.*, les identificateurs infographiques des procédures permettant de construire les relations.

Pour l'analyse linguistique, le système dispose d'un *lexique* qui est une collection de variables. Chacune de celles-ci est un *item* lexical auquel est affectée une définition (catégorie grammaticale ou sémantique, radical, ...). On aura par exemple : *voitures* = *entité*, *radical* : *voiture*, ou bien *roulent-dans* = *relation*, *radical* : *rouler-dans*, ou bien encore *trois* = *adjectif numéral*, *quantité* : 3, où *voitures*, *roulent-dans* et *trois* sont des variables et où la valeur qui leur est affectée est leur définition.

Exemple de traitement d'une phrase :

phr. 2• La mairie est à gauche de l'église.

Le système prend chaque mot l'un après l'autre, du début à la fin de la phrase. Si un *item* lexical est déjà mémorisé dans le lexique, il l'enrichit de sa définition. Sinon le traitement devient interactif et il demande à l'utilisateur si le mot est un article défini, un adjectif numéral, une relation, etc. Une fois les informations recueillies, il range l'*item* lexical et sa définition dans le lexique et procède à l'enrichissement de la phrase.

phr. 2• enrichie : la (adjectif défini, quantité:1) mairie (entité, radical: mairie) est-a-gauche (relation, radical:etre-a-gauche) l' (adjectif défini, quantité:1) église (entité, radical: église) .

Pour effectuer la relation de positionnement, le système doit avoir en mémoire les formes à positionner et de référence ainsi que la procédure infographique réalisant la relation. Le système doit donc d'abord vérifier si les *référents infographiques* —formes et procédure de positionnement— des entités et de la relation linguistiques existent. Il le fait par l'entremise des dictionnaires de correspondances. Pour chaque entité et pour la relation, le système cherche si leurs radicaux figurent dans les dictionnaires. La mairie apparaissant pour la première fois dans le texte, le radical *mairie* ne figure pas dans *DictEnt*, le dictionnaire des correspondances des entités. Le système passe donc la main à l'utilisateur qui crée une nouvelle forme à laquelle il attribue le nom *Mairiel*, en tant que première occurrence des objets infographiques appartenant à la classe *Mairie*. L'utilisateur repasse alors la main au système. Celui-ci range la chaîne de caractères *mairie* suivie de la chaîne de caractères *mairiel* dans *DictEnt* puis, le numéro de la forme de nom *mairiel* étant 3, il affecte ce nombre à la variable *Entité1* du moule : *Entité1*=3. En ce qui concerne l'église, le système trouve *église* associé à *églisel* dans *DictEnt* puisqu'un objet infographique de nom *églisel* a été créé pour la *phr. 1•* précédente [cf. p. 10]. Considérant que les occurrences lexicales d'*église* des *phr. 1•* et *phr. 2•* ont le même référent infographique, le système récupère le numéro de la forme, de nom *églisel*, et l'affecte à *Entité2* : *Entité2*=2. Le radical de la relation, *etre-a-gauche-de*, ne figurant pas dans *DictRel*, le

dictionnaire des correspondances des relations, le système passe la main à l'utilisateur qui élabore une procédure de nom *macetragauchede* permettant de placer un objet X à gauche d'un objet Y par rapport à un observateur. Le système range ensuite la chaîne de caractères *etre-a-gauche-de* suivie de la chaîne de caractères *macetragauchede* dans *DictRel*, puis il lance la procédure *macetragauchede* en instanciant ses paramètres par 3 et 1, les numéros des formes *Mairie1* et *Eglise1*. Dans la scène, visualisée à l'écran, la mairie est alors placée à gauche de l'église.

Problèmes posés par la modélisation

Les deux types de modélisation, linguistique et iconique, ainsi que leur articulation ont soulevé divers problèmes. On notera que la modélisation linguistique est restée rudimentaire du fait que nous ne disposions pas d'un analyseur sémantico-syntaxique, mais uniquement des commandes de traitement de chaînes de caractères d'IKOGRAPH. A ce stade de notre étude, nous n'avons pas cherché à coupler un analyseur au système de synthèse d'image, préférant nous consacrer aux problèmes de modélisation dans un premier temps. Diverses équipes, travaillant dans le domaine de l'intelligence artificielle, ont mis au point des analyseurs performants fonctionnant en particulier sur des dialogues. Mieux vaut donc, ultérieurement, utiliser un analyseur existant que d'essayer d'en réaliser un nouveau.

La modélisation linguistique

Les principales difficultés auxquelles nous nous sommes heurtés sont la quantification indéterminée et la recherche des référents. Lorsque la quantité d'une entité n'est pas précisable à l'aide du déterminant du substantif, comme dans "des maisons" [*phr. 5*•, p. 10] le nombre 100 est choisi; il indique que l'on peut utiliser jusqu'à 100 occurrences de l'entité dans la relation de positionnement.

Le problème du référent s'est posé pour chaque *item* lexical désignant une entité ou une relation et également de façon plus compliquée pour des entités sous-entendues. Deux sortes de référent sont en jeu : le *référent interne* [Greimas & Courtès, 1979, p. 311-312] de nature linguistique et le *référent externe* de nature infographique. On remarquera que dans le contexte d'une interface texte-image de synthèse la question de la vériconditionnalité n'est pas pertinente. Il n'y a pas lieu de se demander si un *item* lexical réfère ou non à un monde réel extérieur ou à un monde possible. La situation est nette : les *items* lexicaux réfèrent à des

phénomènes infographiques. Ainsi, dans notre cas, un *item* lexical du type entité réfère à un objet infographique, et un *item* du type *relation* à une opération infographique qui réalise une relation spatiale. Le système ne disposant pas d'un analyseur sémantico-syntaxique capable de trouver des référents sur des indices strictement linguistiques, pas plus que de mécanismes de raisonnement sur des connaissances de sens commun, la recherche des référents se ramène à la recherche des référents infographiques *via* les dénominations verbales des relations et des classes d'entités.

On a vu plus haut, à propos des dictionnaires de correspondances, quelques exemples de rappel d'objets ou de procédures infographiques à partir des radicaux verbaux désignant des classes d'entités ou des relations. Des procédures *ad hoc* ont été implantées pour traiter les cas où le nombre des référents est pluriel, comme dans trois immeubles (*phr. 10•*, p. 10), et où une entité est sous-entendue. Par exemple, pour la *phr. 6•*, *A gauche [de la rue] il y a trois immeubles*, où *la rue* est sous-entendue, le système revient au moule de la phrase précédente conservée dans sa mémoire et demande si la seconde entité de celui-ci, *Rue1*, est la valeur à affecter à *Entité2* de la *phr. 10•*. L'utilisateur répondant par l'affirmative, il réalise l'affectation : *Entité2 (de la phr. 10•) = Rue1*.

Il est évident que la disposition d'un analyseur sémantico-syntaxique permettrait de complexifier la phase de modélisation linguistique et de traiter les difficultés rencontrées — ainsi que d'autres problèmes que nous avons éludés tels que les anaphores et les synonymes — en mettant en œuvre des connaissances linguistiques et de sens commun ainsi que divers mécanismes de raisonnement [Sabah, 1989, p. 187-316]. Considérons, par exemple, les trois dernières phrases : *Une voiture est sur la place. Une autre voiture est au centre de la place. Trois voitures roulent dans la rue.* Des objets de type *voiture* apparaissant pour la première fois dans le texte, sous cette appellation ou sous une autre, il semble que les déterminants *une* puis *une autre* puis *trois*, non précédés par un article défini, suffisent linguistiquement pour savoir que la seconde voiture n'est pas la première et que les trois dernières ne sont pas les deux premières. Un raisonnement de sens commun est aussi possible qui établirait : 'les deux premières voitures sont sur la place, donc elles ne peuvent pas être deux des trois voitures qui roulent dans la rue.'

La modélisation spatio-visuelle

Le problème pour l'ingénieur, qui peut être aussi le concepteur de la scène, est de traduire une image mentale tridimensionnelle en des descriptions formalisées d'opérations à effectuer d'une part pour construire des objets et d'autre part des relations de positionnement. Deux types de représentations mentales imagées sont donc impliqués. Les uns

de nature descriptive découlent du texte de départ tandis que les autres, de nature opérative, sont à transcrire en des textes rédigés dans le langage formalisé du système de synthèse d'image. Les images mentales du premier type sont des "modèles visuo-spatiaux (...) construits à partir de messages linguistiques" [Denis, 1989, p.211]. Elles subissent les influences de différents facteurs : la tâche qui consiste à construire une scène fixe de nature schématique, le contexte imagé attaché au contexte linguistique, le type de modélisation géométrique mis en oeuvre par le système de synthèse d'image. Le fonctionnement cognitif est économique, en ce sens qu'il se situe à des niveaux de détail variés selon les besoins du moment [Coulon & Kayser, 1982]. Le but du concepteur étant de positionner des objets schématiques les uns par rapport aux autres, les représentations mentales restent assez vagues, sauf en ce qui concerne les éléments de la scène pertinents pour la relation de positionnement. Dans la *phr. 5*•, *Des maisons bordent le côté droit de la rue*, les traits saillants sont le bord de la rue et les bases des façades des maisons. La rue des *phr. 4*• et *5* peut se réduire à une schématisation où le trottoir n'apparaît pas. Ce n'est qu'à la *phr. 11*•, où le trottoir est mentionné explicitement, qu'un trottoir est ajouté à la rue [cf. texte, p. 10]. Des formes géométriques de base, faciles à exécuter par le système —parallélépipèdes rectangles pour les bâtiments, ruban rectangulaire pour la rue, etc.— sont prédominantes. L'intervention du contexte imagé provoque des ajustements. L'église de la *phr. 1*• et la voiture de la *phr. 9*• sont décrites comme étant toutes les deux au centre de la place. Le centre est considéré comme un point pour la première phrase, puis comme une zone pour la neuvième : la voiture est alors située près de l'église dans la région centrale de la place

La prise en compte anticipée du type de modélisation géométrique implanté dans le système provoque une géométrisation des images mentales descriptives. La géométrisation est faite mentalement et en s'aidant de croquis et de manipulations de papiers découpés ou d'objets parallélépipédiques. Elle s'est effectuée en trois puis en deux dimensions. En effet tous les volumes étant des parallélépipèdes rectangles ou ayant une enveloppe de cette forme, il suffit de raisonner en deux dimensions, sur les fonds rectangulaires situés dans le même plan pour positionner correctement les objets les uns par rapport aux autres.

Les images mentales *opératives* interviennent une fois les représentations géométriques des relations définies, car il s'agit alors d'imaginer quelles sont les opérations géométriques, inspirées des commandes du système, qui doivent être réalisées pour positionner les représentations géométriques des fonds les uns par rapport aux autres. Ces opérations sont essentiellement la translation et la rotation auxquelles il faut ajouter la duplication. Cependant une géométrisation de type euclidien ne suffit pas car le système opère dans un univers *vectorisé*, relevant de la géométrie analytique. Tout doit être orienté : les arêtes des facettes deviennent des vecteurs, les angles sont parcourus dans le sens

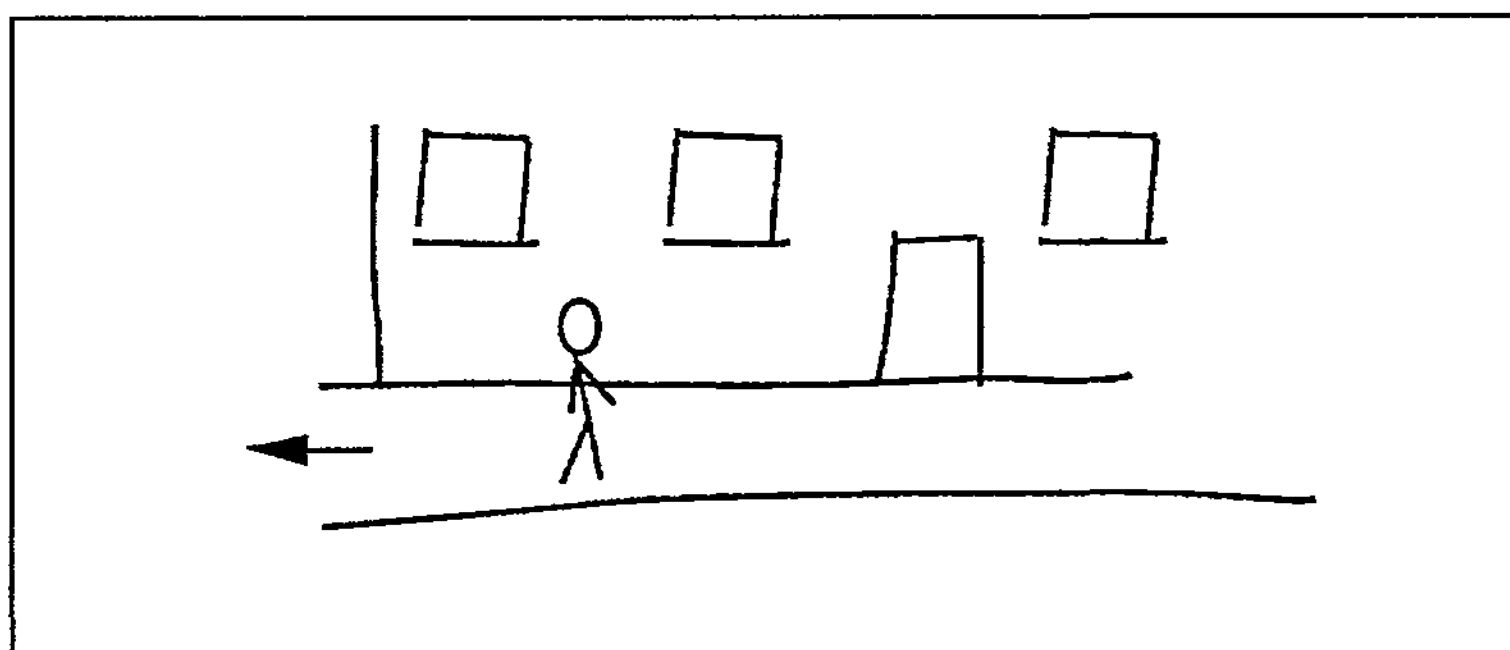
positif ou négatif, les translations s'opèrent le long de vecteurs, etc. Tout est mesuré également en fonction des trois axes orthogonaux OX, OY, OZ. On notera que les différents paliers du processus de modélisation spatio-visuelle ne sont pas aussi nettement séparés et ordonnés dans la réalité que dans la présentation qui vient d'en être faite. Au contraire, ils s'entremêlent souvent. Il arrive même, dans certains cas, que les phases intermédiaires soient court-circuitées et que l'on passe quasi directement de l'énoncé verbal décrivant une relation à la définition de l'image opérative vectorisée correspondante et même à la rédaction d'une procédure.

— Exemple de modélisation spatio-visuelle

phr. 11• : Le piéton marche sur le trottoir de droite.

Image mentale descriptive.

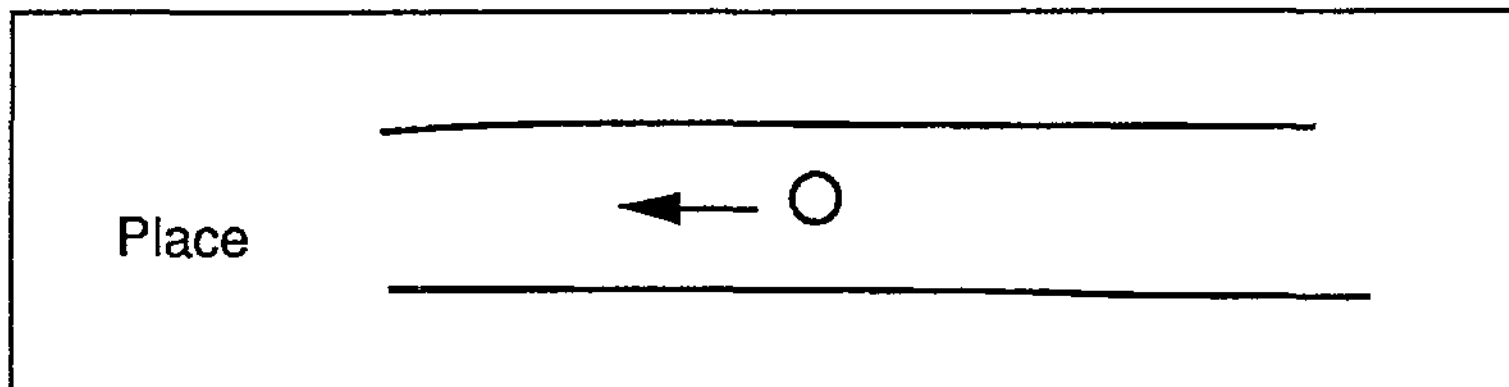
La scène à construire étant fixe, l'action "marcher sur" est transformée en un état. L'aspect duratif du verbe semble avoir eu une influence, car le piéton est imaginé dans une zone intermédiaire du trottoir et non au début ou à la fin de celui-ci. Il en est de même du contexte imagé lié au contexte linguistique que constitue la *phr. 4•* précédente, *La rue arrive à la place.* La rue, peut-être sous l'effet du dynamisme du verbe "arriver à", est ressentie comme orientée vers la place. Le piéton, sur le trottoir lié à la rue, apparaît également comme tourné vers la place.



— Géométrisation de l'image mentale descriptive

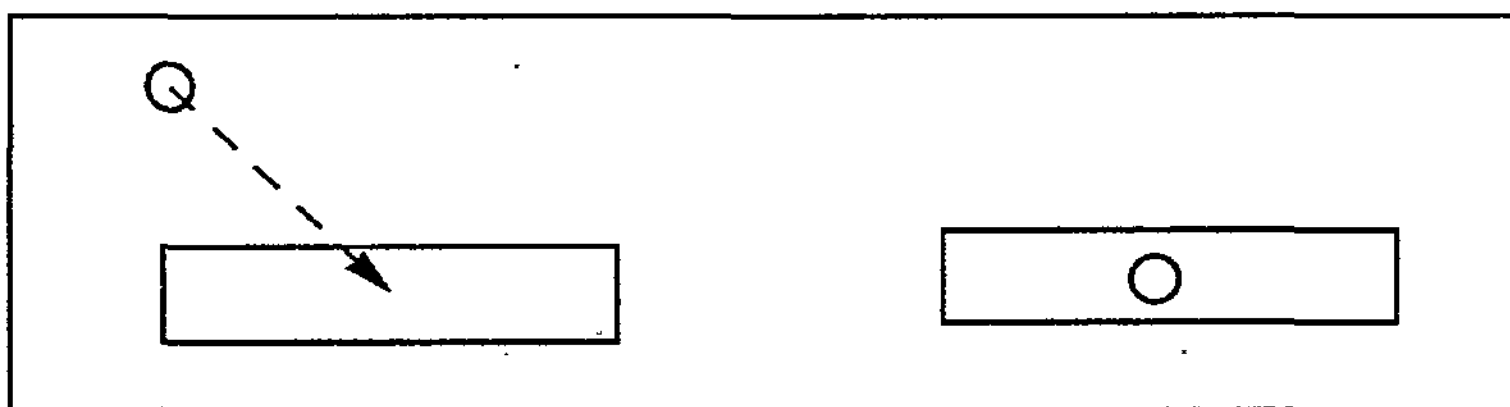
L'univers tridimensionnel, et mobile pour le piéton, est ramené à un univers bidimensionnel totalement fixe. La base du piéton est réduite à un cercle et le trottoir à un ruban dans lequel le cercle est inclus dans une zone intermédiaire. La géométrisation s'accompagne donc d'une forte

désémantisation qui n'est pas complète cependant, car le piéton est toujours vu comme regardant vers la place.



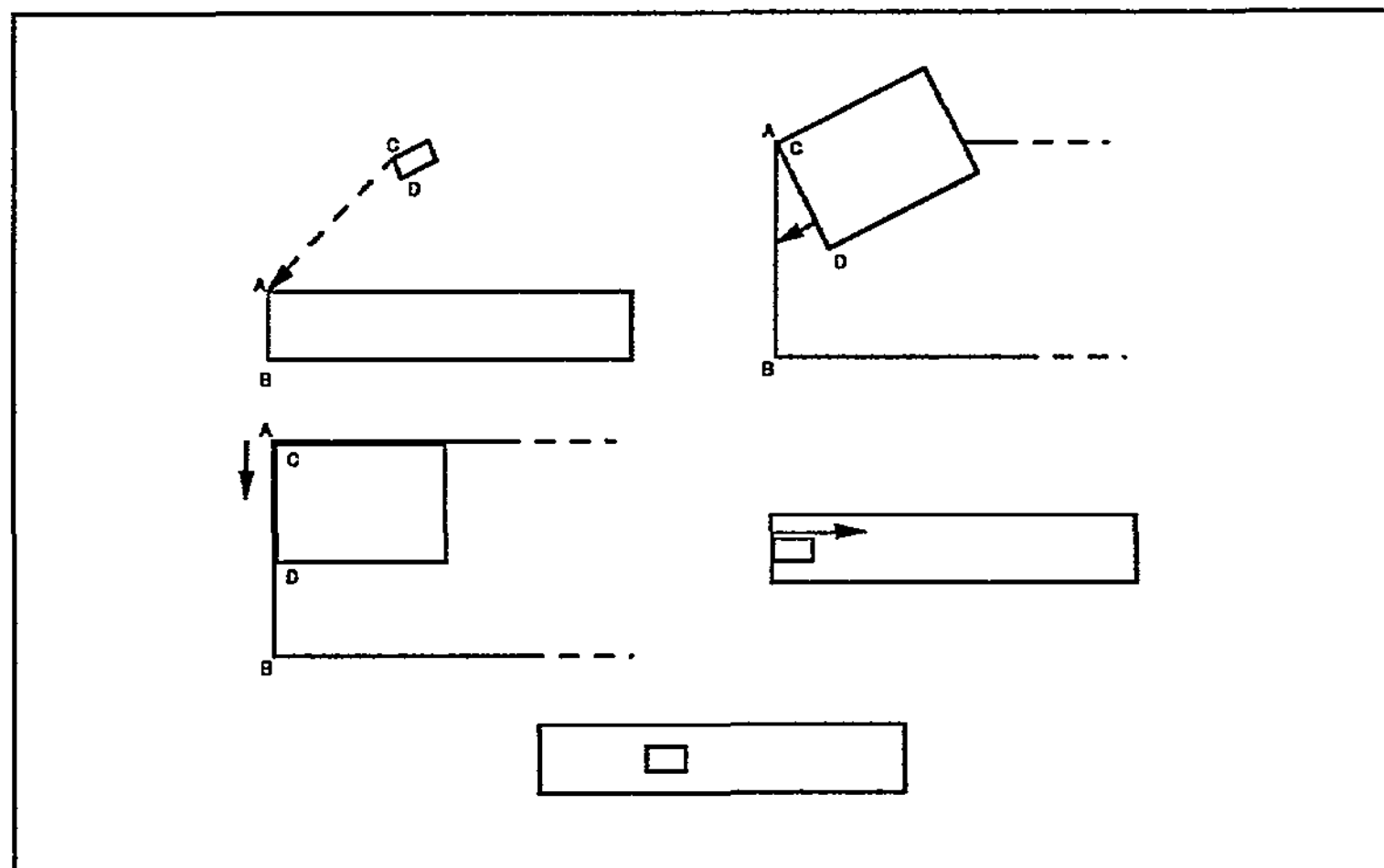
— Géométrisation des images mentales opératives

Connaissant les commandes disponibles sur le système, une forme rectangulaire est choisie pour le trottoir. L'opération sélectionnée pour amener la base du piéton à l'intérieur du trottoir, est une translation qui fait coïncider le centre de gravité du cercle avec celui du rectangle. Le contexte opératif, constitué par les opérations définies pour des phrases précédentes, se manifeste à travers le choix de la translation et des centres de gravité. En effet, pour les *phr. 1•* et *phr. 9•*, *L'église est au centre de la place* et *Une voiture est au centre de la place*, une translation fait coïncider le centre de gravité du fond de l'objet à positionner avec celui du fond de l'objet de référence. La voiture est déplacée ensuite à côté de l'église pour éviter qu'elle ne soit à l'intérieur du bâtiment... L'image mentale opérative ne présente plus guère de trace sémantique de l'univers de sens commun à ce stade ; le piéton, en particulier, perd son orientation vers la place et se réduit à un cercle à inclure dans un rectangle.



— Formalisation des images opératives

L'influence du contexte opératif est déterminante et provoque la transformation du cercle de base du piéton en un rectangle, et la décomposition de l'opération unique de translation en trois translations et une rotation. Une procédure ayant été créée pour positionner le fond rectangulaire de la première voiture dans la *phr. 10•*, *Trois voitures roulent dans la rue*, elle est réutilisée pour placer la base du piéton à l'intérieur du trottoir.



Descriptions sémantique et imagée d'un énoncé verbal

Le problème central soulevé par la transcription automatique d'un texte en une scène est celui des relations à établir entre les descriptions sémantique et imagée de la scène. Nous disons imagée plutôt que géométrique pour le second type de description car la modélisation géométrique d'une scène n'est qu'une possibilité parmi d'autres de représentation. Un utilisateur a tout loisir de jouer aussi sur les couleurs, les lumières, les textures, les transparences, etc. pour modifier l'aspect de la scène et la rendre, par exemple, de façon *réaliste*... La question des rapports entre descriptions sémantique et imagée amène à s'interroger d'une part sur la capacité d'une langue à traiter du sensible et d'autre part sur le partage des compétences entre sémiotique (et linguistique) et psychologie cognitive dans le cadre d'une collaboration interdisciplinaire.

En psychologie cognitive, les relations entre énoncés verbaux et images mentales spatio-visuelles attachées à ceux-ci font partie du domaine clé que sont les "représentations mentales" [Denis, 1989, chap.1 ; Denhière & Baudet, 1992, chap.1]. La représentation mentale est un phénomène psychologique qui est soit une activité soit le produit de cette activité. En tant que produit, elle est "quelque chose qui tient pour quelque chose d'autre, c'est un modèle, un représentant de la chose représentée"

[Denhière & Baudet, 1992, p. 36]. Les images mentales, dont celles qui sont de nature visuelle, constituent un type de représentations apparentées aux percepts [Denis, 1989, p. 64-74]. Un autre type de représentation important est la représentation qualifiée de propositionnelle qui concerne en particulier la sémantique des énoncés verbaux et qui donne lieu à une modélisation en termes de prédicats et d'arguments, d'inspiration logique [Denhière et Baudet, 1992, p. 41-54]. Si la représentation de type propositionnel est une composante essentielle dans le traitement du texte, elle s'accompagne souvent de représentations imagées plus ou moins élaborées, et plus ou moins abstraites [Denis, 1989, chap. 4 à 7]. Celles-ci ont des "règles d'accès (...) similaires à celles des événements perceptifs", aussi permettent-elles de "dériver des inférences sans devoir recourir, par exemple, à des calculs fondés sur la logique formelle" [Denis, 1989, p. 213]. Les images mentales ne sont pas, cependant, le seul mode de représentation des phénomènes d'ordre perceptif décrits par un texte. Les modèles de signification textuels peuvent inclure des descriptions propositionnelles de phénomènes de nature spatio-temporelle. Ces descriptions mettent en œuvre des "catégories conceptuelles" telles que celles "d'objet (ou individu), d'état, d'événement, d'action et de relations (causales, temporelles, topologiques, méréologiques, etc.)" [Denhière & Baudet, 1992, p. 163, 168-189].

Les sémanticiens qui s'intéressent à "la partie figurable de la signification" [Denis, 1989, p. 129] ou plus généralement à la "perception sémantique" [Rastier, 1991, chap.4], considèrent également que certains énoncés verbaux donnent lieu à une représentation imagée qui n'est pas de même nature que la partie de la signification qui est descriptible par des propositions ou par des structures sémiques [Herskovits, 1986, p. 100-103 ; Rastier, 1991, p. 207]. S'ils sont d'accord sur la nature cognitive de la représentation imagée, il n'en est pas de même pour celle de la représentation linguistique. En effet, les sémanticiens cognitivistes, tout comme les psychologues cognitivistes, ont tendance à assimiler le signifié au concept [Rastier, 1991, chap. 3].

Pour F. Rastier, la frontière est nette entre les compétences du sémanticien et du psychologue cognitiviste en ce qui concerne le traitement des énoncés verbaux suscitant des images mentales, autrement dit des "simulacres multimodaux" relevant des différents registres sensoriels. Pour lui, "les images mentales sont des corrélats psychologiques des signifiés linguistiques" et elles ne relèvent donc pas du sémanticien [Rastier, 1991, p. 207]. Par contre, il appartient à celui-ci de décrire les sèmes spatio-temporels qui entrent dans les signifiés. Ainsi F. Rastier définit /rondeur/ comme un sème définitoire du lexème *tonneau* au

sein de la classe sémantique de celui-ci [Rastier, 1991, p. 214]. Ou encore, analysant le début du vers d'Apollinaire, "Bergère ô tour Eiffel....." —où les expressions verbales *bergère* et *tour Eiffel* sont réunies linguistiquement— il considère que le sème /verticalité/ entre dans le sémème *bergère* du fait du lien de celui-ci avec *tour Eiffel*, tandis que les deux sémèmes s'opposent par les sèmes /petit/ et /grand/ dans ce contexte précis [Rastier, 1989, p. 252, 256].

Une description de type propositionnel ou sémique étant incapable de rendre compte de l'aspect sensible d'une représentation, il est nécessaire de trouver un formalisme adapté à la modélisation de celui-ci. La théorie de la vision développée par David Marr pour la compréhension automatique de l'image apparaît comme une source de modèles intéressante pour traiter les images mentales tridimensionnelles accompagnant les signifiés linguistiques [Vaina, 1983 ; Herskovits, 1986, p. 100-105 ; Jackendoff, 1987]. Etant donnée une scène visualisée par une image, D. Marr part de la mosaïque de points plus ou moins gris qui constitue celle-ci, apparentée à l'image rétinienne [Marr, 1982]. L'extraction des éléments pertinents et leur interprétation s'effectue graduellement. Au plus bas niveau, bidimensionnel, sont détectés les points caractéristiques, c'est-à-dire les points où se produit un changement d'intensité lumineuse. Ces points sont considérés comme des événements *significatifs* dans la scène en ce sens qu'ils indiquent souvent des bords d'objets ou d'ombres, des ruptures ou des graduations dans les orientations des surfaces, des plans de profondeur différents, etc. [Brady, 1982, p.16]. Puis sont définies les lignes ouvertes ou fermées formées par ces points. A un niveau intermédiaire, en ronde bosse si l'on peut dire, les orientations des surfaces visibles et les ruptures entre orientations sont déterminées —les surfaces étant délimitées par les contours et coïncidant avec les régions, identifiées au niveau inférieur. Enfin au dernier niveau, tridimensionnel, les volumes des objets sont inférés des surfaces visibles. Un processus de reconnaissance de formes est complet si "l'interprétation" est conjuguée à "la perception visuelle" [Faure, 1988, p. 175-177] ; les formes sont alors identifiées comme des objets connus dans un domaine de connaissances particulier ou bien dans le monde du sens commun.

On remarquera que les modèles de D. Marr —et, de façon générale, en reconnaissance des formes et en compréhension automatique de l'image— ne sont pas sans évoquer les approches de la sémiotique plastique (J.M. Floch *in* [Greimas & Courtès, 1986, p. 169-170]). En effet l'analyse du plan de l'expression de l'image bidimensionnelle, et éventuellement de la scène tridimensionnelle qu'elle représente, est clairement distinguée de l'interprétation, c'est-à-dire de la mise en rapport

du plan de l'expression avec le plan du contenu. L'analyse du signifiant comporte plusieurs paliers qui, du plus élémentaire au plus complexe, sont ceux :

- des traits descriptifs relevant de "catégories visuelles" élémentaires se rapportant au chromatisme, au graphisme, à la lumière, à la profondeur, etc. Exemple : les catégories *droit vs courbe*, *long vs court*, *segmenté vs non segmenté*, utilisées pour caractériser les lignes élémentaires apparaissant dans un tableau de Kandinsky [Floch, 1985, p. 64] ;

- des "figures de l'expression" qui sont des combinaisons de traits élémentaires. Exemple : les "noeuds", les "hachures", les "arborescences" repérés dans le même tableau de Kandinsky [Floch, 1985, p. 65].

Avec les "formants" commence l'interprétation. Les formants sont des parties du plan de l'expression auxquels est attachée une signification ; ils intègrent des figures de l'expression. Ainsi en est-il des formants de *l'arc-en-ciel*, des *cavaliers* et de la "pente hérissée" dans des tableaux de Kandinsky [Floch, 1985, p. 46-49], et de ceux du *tuyau de poêle* et de la *niche* dans des bandes dessinées de Benjamin Rabier [Courtès, 1992, p. 54].

J. Courtès souligne que "l'interprétation des données visuelles" dépend d'un savoir qui est lié à un "contexte socio-culturel déterminé". Par ailleurs, un même formant peut "recevoir— dans des contextes différents— diverses interprétations sémantiques" [Courtès, 1992, p. 36]. Dans un contexte visuel autre, le dessin du tuyau, par exemple, aurait pu être identifié comme une planche ou un barreau, et celui de la niche comme une maison [Courtès, 1992, p. 54].

Alors qu'en reconnaissance des formes et en compréhension automatique de l'image, la dimension narrative de l'image n'est pratiquement pas abordée, en sémiotique plastique cet aspect est communément traité. Les formants jouent alors des rôles dans des histoires ou des fragments d'histoires traités à des niveaux plus ou moins concrets.

Types de description à l'œuvre dans la transcription automatique d'un texte en une scène

Les incursions que l'on vient d'effectuer dans différentes sciences cognitives amènent à distinguer trois types de description à l'œuvre dans la transcription automatique d'un texte en une scène. Le premier, que nous qualifierons de *physique*, mais qui est uniquement de nature géométrico-algébrique dans notre cas, rend compte de la structure spatio-visuelle des

représentations imagées associées aux signifiés linguistiques. Pour des traitements de scènes plus élaborées, il inclut des modèles physiques ou mécaniques, par exemple, pour traiter les couleurs, les lumières, les déplacements ou d'autres facteurs. Le deuxième type de description, d'ordre *sémique*, modélise les signifiés tels qu'ils se constituent les uns par rapport aux autres au sein du texte. Le qualificatif "sémique" est utilisé ici, de préférence à "propositionnel", car il nous semble plus général. L'analyse sémique des signifiés en sémiotique n'est pas formalisée généralement, cependant elle pourrait l'être et donner lieu à une représentation et à des calculs propositionnels comme le montre la réflexion de F. Rastier sur la représentation du signifié lexical à l'aide des réseaux sémantiques utilisés en intelligence artificielle [Rastier, 1991, p. 140-161]. Les traits de la description sémique caractérisent des entités —individu, classe, tout ou partie d'un tout— et les propriétés intrinsèques et relatives de celles-ci. Dans la description sémique entrent des traits aussi bien figuratifs que non figuratifs. On remarquera que ceci est conforme au traitement sémantique préconisé par F. Rastier qui considère, par exemple, que /rondeur/ est un sème parmi d'autres du sémème "tonneau".

Le troisième type de description, que nous qualifierons de *phémique*, concerne les représentations imagées associées aux signifiés linguistiques. Rappelons que le phème est l'équivalent du sème en ce qui concerne le plan de l'expression. Unité minimale du plan de l'expression, il appartient au métalangage théorique et "ne relève d'aucune substance (d'aucune "réalité")" [Greimas et Courtés, 1979, p. 276-277]. Les traits issus des catégories visuelles, cités plus haut à propos des tableaux de Kandinsky et de la bande de Benjamin Rabier, sont des phèmes. C'est le décrochement de la "réalité" qui fait la différence radicale entre descriptions phémique et physique. Encore faut-il préciser en quoi consiste cette "réalité" mise à juste titre entre guillemets. Il ne s'agit pas d'une réalité objective, mais d'une réalité construite. Le percept ou l'image mentale d'une scène, une maquette virtuelle tridimensionnelle, un dessin, un organigramme, sont les produits du "schématisme" [Metz, 1974, p. 142]. Celui-ci s'apparente à la catégorisation, phénomène mental étudié en psychologie cognitive [Dubois, 1991 ; Denis, 1989 ; Denhière & Baudet, 1992]. Les schémas, structures de traits pertinents, sont construits à partir de l'expérience vécue et sont dépendants du contexte socio-culturel. Ils permettent à l'individu d'interpréter son environnement —monde réel ou mondes des signes— et d'agir sur lui. Dans le cas d'une image figurative, "l'impression référentielle", ou encore l'impression que l'on a affaire à un monde sensible, est issue de transformations, réglées par des conventions, et qui s'appliquent à certains des traits pertinents du schéma perceptuel. Ainsi en est-il, par exemple, de la silhouette d'un cheval dessinée d'une seule ligne noire [Eco, 1976, p. 194].

La description phémique tout comme la description sémique permet de faire des raisonnements logiques, implantables informatiquement. Ainsi en est-il des descriptions mémorisées sous la forme de réseaux sémantiques dans le système ANALOGY couplé à un système de compréhension de l'image [Winston *et al.*, 1984, p. 135] :

Exemple :

- Description de la classe TASSE :

TASSE : est un OBJET
est un RECIPIENT OUVERT
a pour propriété d'être STABLE
a pour propriété d'être PORTABLE

- Description d'une valise particulière:

Valise : a une POIGNEE
a pour propriété d'être LEGER
a pour propriété d'être PORTABLE
a pour propriété d'être SAISSISSABLE
est saisissable parce qu'elle est légère et qu'elle a une poignée
est portable parce qu'elle est légère

- Description d'un objet particulier qui a été reconnu comme une tasse :

Tasse : Traits de l'objet identifiés par le système de compréhension de l'image
a une POIGNEE
a un FOND PLAT
a un CORPS PETIT
a une CONCAVITE TOURNEE VERS LE HAUT
Trait ajouté
a pour propriété d'être LEGER
Traits caractéristiques de la classe des tasses
est un OBJET
est un RECIPIENT OUVERT
a pour propriété d'être STABLE
a pour propriété d'être PORTABLE

Raisonnant par analogie avec la valise qui, comme la tasse, a une poignée et est légère, le système ajoute les points suivants à la description de la Tasse :

Tasse : a pour propriété d'être SAISSISSABLE
est saisissable parce qu'elle est légère et qu'elle a une poignée.

Cet exemple montre qu'il n'y a pas de différence entre descriptions phémique et sémique en ce qui concerne la description de scènes. Les descriptions ci-dessus pourraient très bien être issues d'énoncés verbaux. Les descriptions sémiques tout comme les descriptions phémiques peuvent donner lieu à des raisonnements spatio-temporels. Ainsi dans le cas de notre texte, si un utilisateur demandait à un système de compréhension automatique du langage naturel : "Où est la mairie?", le système répondrait sans peine "A gauche de l'église", puisqu'il dispose de cette information sous une forme verbale. A la question "L'église est-elle à droite de la mairie?", il pourrait également répondre "Oui" s'il possédait la règle d'inférence "Si un objet X est à gauche d'un objet Y alors l'objet Y est à droite de l'objet X". Par contre d'autres raisonnements sont impossibles sans recours à la description physique. Si l'on demandait à un système "Combien d'arbres y a-t-il autour de la place?" ou bien "Peut-on voir la rue de derrière la place?", par exemple, les informations fournies par le texte seraient insuffisantes. En effet, dans le texte, le nombre d'arbres est indéterminé ; le texte n'indique pas non plus, directement ou indirectement, où est l'arrière de l'église et où est la rue par rapport à l'église. Pour répondre aux deux questions, il est nécessaire de se reporter à la description physique de la maquette virtuelle de la scène et de raisonner sur elle [voir Annexe 2, p. 29]. Dans la maquette, en effet, les arbres entourant la place sont dénombrés, et la rue est invisible de derrière l'église car elle débouche face à ce bâtiment. Ces observations sur l'incomplétude des informations fournies par le texte, vont dans le sens de la recherche menée par C. Faure [Faure, 1988, p. 194-196] sur les rapports verbaux d'experts analysant des images. Les descriptions verbales se sont avérées "incomplètes", "qualitatives et floues", et "contradictaires". Inversement, la description verbale peut contribuer à enrichir la description physique d'entités et de propriétés qui, bien que spatio-visuelles, ne sont pas évidentes dans "l'apparence visuelle". Ainsi en est-il des relations de partie à tout, des systèmes d'axes coordonnés, des régions déterminées par les axes des objets, des trajectoires suivies par les objets en mouvement [Jackendoff, 1987, p. 110-111].

Conclusion

Au début de notre expérience, nous prévoyions d'avoir affaire à deux sortes de représentations dans la transcription automatique d'un texte en une scène tridimensionnelle, du moins au stade de l'informatisation : une représentation linguistique du contenu sémantique du texte et une représentation géométrico-algébrique de la scène évoquée par celui-ci. Une réflexion plus approfondie, dépassant le domaine restreint du texte, de

sa transcription schématique et des moyens limités du système en matière de compréhension automatique du langage naturel et de l'image, nous a conduits à distinguer une troisième sorte de représentation, que nous avons qualifiée de phémique, et qui traite de la scène spatialisée sous une forme identique à celle de la description sémique. La description phémique peut contribuer à enrichir la description physique comme peut le faire la description sémique. Les raisonnements autorisés par les descriptions sémique et phémique d'une part et description physique d'autre part sont en relation de complémentarité. Ces observations contribuent à éclaircir les rapports entre partie figurable et partie sémantique de la signification. Elles n'auraient pas pu être faites sans l'effort d'explicitation imposée par la nécessité d'implanter informatiquement les modèles élaborés.

Nous remercions Claudie Faure de sa lecture attentive
et de ses observations pertinentes.

Références bibliographiques

BRADY (M.)

1982, "Computational approaches to image understanding", *Computing surveys*, vol. 14, n°1, mars.

CLAYSEN (D.), LOBSTEIN (D.), ZEITOUN (J.)

1987, *Les Nouvelles images : introduction à l'image informatique*, Paris, Dunod.

COULON (D.) & KAYSER (D.)

1982, "La Compréhension : un processus à profondeur variable", *Bulletin de psychologie*, t. 35, n°356, p. 815-823.

COURTÉS (J.)

1992, "Du signifié au signifiant : étude d'une bande dessinée de Benjamin Rabier", *Nouveaux actes sémiotiques*, n°21-22.

DENHIÈRE (G.) & BAUDET (S.)

1992, *Lecture, compréhension de texte et science cognitive*, Paris, PUF.

DENIS (M.)

1989, *Image et cognition*, Paris, PUF.

DUBOIS (D.)

1991, *Sémantique et cognition : catégories, prototypes, typicalité*, Paris, Editions du CNRS.

ECO (U.)

1978, "Pour une reformulation du concept de signe iconique : les modes de production sémiotique", *Communications*, n°29, p. 141-191.

FAURE (C.)

1988, "Reconnaissance et description de formes", "Perception visuelle", *Intellectica*, n°5, p. 175-206.

FLOCH (J.M.)

1985, *Petites mythologies de l'œil et de l'esprit : pour une sémiotique plastique*, Paris-Amsterdam, Hatje-Benjamins.

GREIMAS (A.J.) & COURTÉS (J.)

1979, 1986, *Sémiotique : dictionnaire raisonné de la théorie du langage*, Paris, Hachette, t. 1, t. 2.

GUENA (F.)

1992, "Un Raisonnement par analogie pour la résolution de problèmes de conception qui requièrent de l'innovation", *Vers une nouvelle génération des systèmes d'aide à la conception. Actes de OI Design'92*, Marrakech, 25-27 janvier 1992, Paris, Europa.

HERSKOVITS (A.)

1986, *Language and spatial cognition*, Cambridge University press.

JACKENDOFF (R.)

1987, "On beyond zebra : the relation of linguistic and visual information", *Cognition*, vol. 26, n°2, p. 89-114.

Manuel d'utilisation d'IKOLIGHT

1989, Pub. de l'Atelier d'informatique de l'Ecole d'architecture de Normandie.

MARR (D.)

1982, *Vision : a computational investigation into the human representation and processing of visual information*, New York, Freeman.

METZ (Ch.)

1977, "Le Perçu et le nommé", *Essais sémiotiques*, Paris, Klincksieck, ch. 6, p. 130-161.

RASTIER (F.)

1989, *Sens et textualité*, Paris, Hachette.

1991, *Sémantique et recherches cognitives*, Paris, PUF.

SABAH (G.)

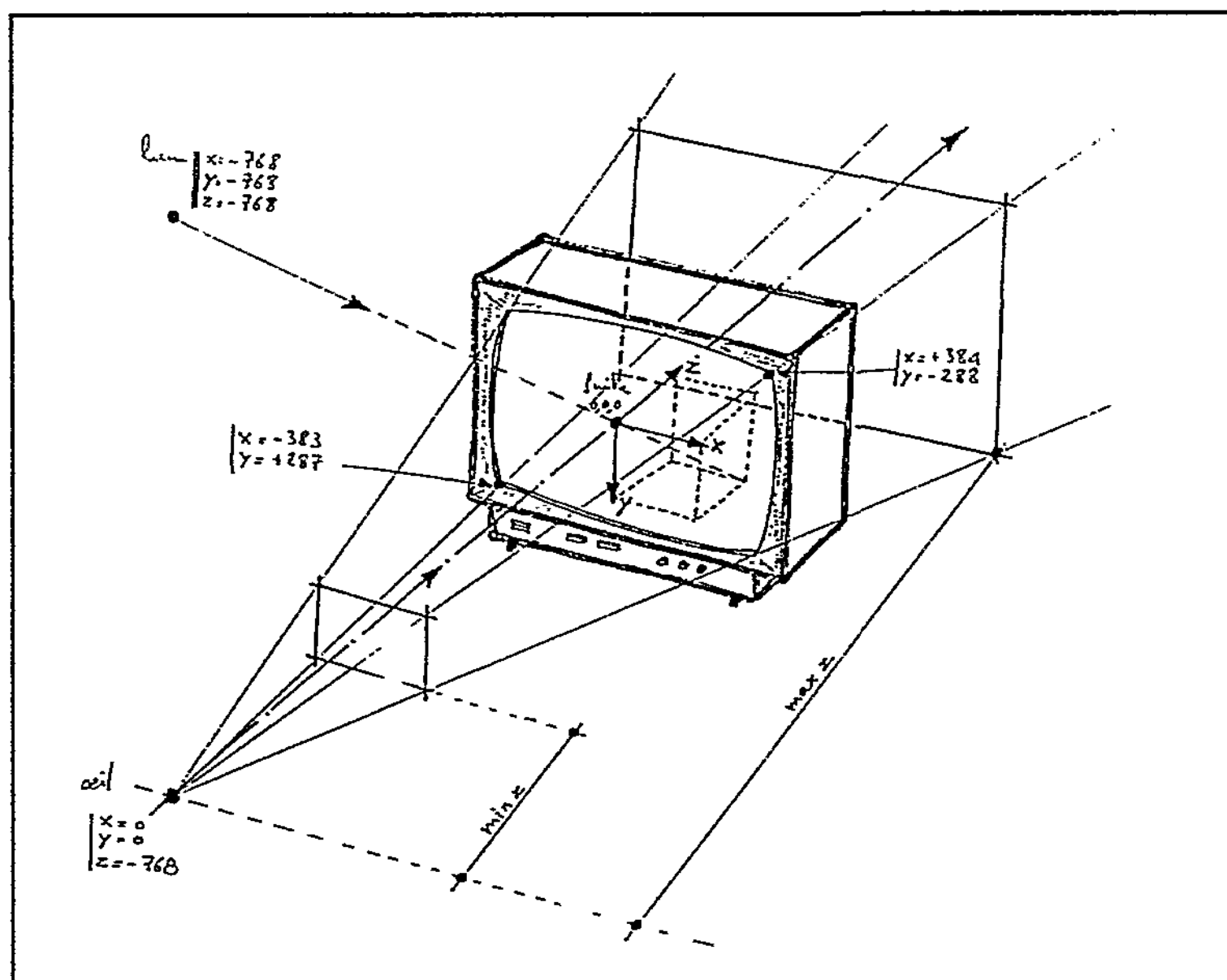
1989, *L'Intelligence artificielle et le langage*, vol. 2 : *Processus de compréhension*, Paris, Hermès.

VAINA (L.)

1983, "From shapes and movements to objects and actions : design constraints on the representation", "Problems of perception", *Synthese*, n°54, p. 3-36.

WINSTON (P.H.) *et al*

1984, "Learning physical descriptions from functional definitions, examples, and precedents", *Robotics research*, ed. by M. Brady and R. Paul, Cambridge, Mass., The MIT press, p. 117-135.



L'univers d'IKOGRAPH

Unité de mesure : le pixel, autrement dit un carreau de la mosaïque qu'est l'image.

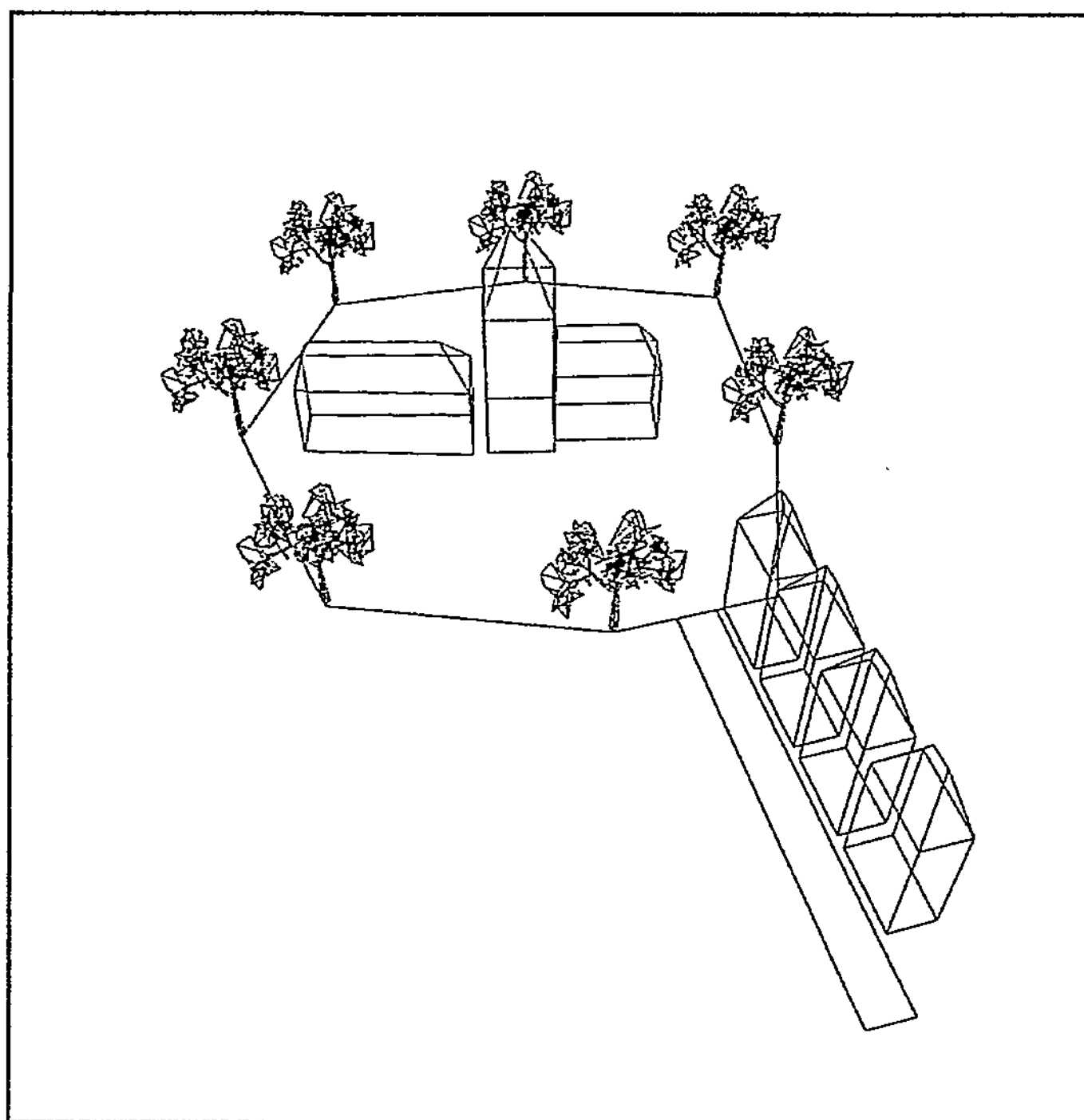
- *Lum* : position par défaut de la lumière qui éclaire la scène.
- *0, 0, 0* : coordonnées de l'origine des axes x, y et z par rapport à laquelle sont construits et positionnés les objets de la scène.
- *Oeil* : position par défaut de l'oeil de l'observateur.
- *Fuite* : point de visée de l'oeil. Ses coordonnées sont 0, 0, 0 par défaut.

En général on choisit le plan xOy comme sol, les formes architecturales s'élevant du sol vers l'observateur, c'est-à-dire de zéro vers -z. Le résultat équivaut à une vue aérienne de la scène. Pour avoir une perspective avec un sol horizontal on change la position de l'oeil, ou bien on fait une rotation du sol autour de l'axe des x.

Exemples d'opérateurs graphiques d'IKOGRAPH

- Création d'objets : boule 50, 10
(Créer une sphère dont le rayon est long de 50 pixels et qui compte 10 tranches).
- Modification d'objets : dilx. for 1 ; coe = 30
(Dilater la forme 1 suivant l'axe des x ; coefficient de dilatation = 30).
- Opérations spatiales sur les objets : dep. for 3 ; poi = 100, 200, 50
(Déplacer la forme jusqu'au point de coordonnées 100, 200, 50).

Annexe 1



Tirage d'une image affichée à l'écran

La scène correspond aux quatre premières phrases :

*L'église est au centre de la place.
La mairie est à gauche de l'église.
La place est entourée d'arbres.
La rue arrive à la place.*

Annexe 2

