



HAL
open science

Représentation et interprétation de connaissances spatiales dans un système de dialogue Homme-Machine

Daniel Schang

► **To cite this version:**

Daniel Schang. Représentation et interprétation de connaissances spatiales dans un système de dialogue Homme-Machine. Informatique [cs]. Université Henri Poincaré - Nancy 1, 1997. Français. NNT : 1997NAN10009 . tel-01754336

HAL Id: tel-01754336

<https://hal.univ-lorraine.fr/tel-01754336>

Submitted on 30 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>



FACULTÉ DES SCIENCES

U.F.R. S.T.M.I.A.
École Doctorale IAE+M
Département de Formation Doctorale en Informatique



THÈSE

présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy I
en Informatique

par Daniel Schang

Représentation et interprétation de connaissances spatiales dans un système de dialogue homme-machine.

Soutenance publique le 13 Janvier 1997 devant la commission d'examen :

Membres du jury :

<i>Président :</i>	M. Jean-Paul HATON	Professeur, Université Henri Poincaré, Nancy I
<i>Rapporteurs :</i>	M. Gérard SABAH M. Wolfgang WAHLSTER	DR au CNRS (LIMSI, Orsay) Professeur, Université de Saarbrücken
<i>Examineurs :</i>	M. Jean-Marie PIERREL M. Laurent ROMARY Mlle. Laure VIEU	Professeur, Université Henri Poincaré, Nancy I CR au CNRS (CRIN, Nancy) CR au CNRS (IRIT, Toulouse)

Résumé

Cette thèse présente un ensemble de recherches dont l'objectif est de contribuer à la mise en œuvre de systèmes de dialogue homme-machine possédant un comportement aussi naturel que possible. Ces travaux portent sur le domaine de la référence aux objets pour des dialogues homme-machine multimodaux. Plus spécifiquement, dans le cadre d'une coopération entre expressions langagières et visualisation graphique d'un espace de désignation, nous nous intéressons aux références spatiales s'exprimant à l'aide de propositions spatiales. La réflexion proposée s'articule autour des quatre axes principaux suivants :

- dans un premier temps, nous fournissons la trame de cette thèse qui se fonde sur le triptyque suivant : dialogue, contexte et principe d'économie cognitive,
- nous cherchons dans un second temps, à élaborer un modèle qui puisse effectuer des références spatiales portant sur un environnement accessible visuellement à un utilisateur,
- le caractère par trop *ad hoc* du modèle précédent nous incite, dans le cadre des sciences cognitives, à élaborer un modèle de représentation de l'espace qui permet de considérer sous un même angle des phénomènes liés au langage, à la représentation des connaissances et au raisonnement spatial et ce, à l'aide de la notion générale de cadre. Cette notion est instanciée pour le problème spécifique des énoncés de référence spatiaux sous la dénomination de cadre de référence dont nous décrivons les tenants et les aboutissants,
- l'aboutissement de ce travail permet de donner certaines pistes en vue de l'intégration du modèle proposé dans un système opérationnel.

Mots-clés: dialogue homme-machine, langage naturel, référence, modélisation spatiale.

Remerciements

Je souhaite remercier vivement les membres du jury : les rapporteurs, Jean-Paul Haton, Gérard Sabah et Wolfgang Wahlster, pour avoir accepté de juger cette thèse et pour les conseils précieux qu'ils m'ont donnés, ainsi que Laure Vieu qui a examiné ce manuscrit avec attention.

Mes co-directeurs de thèse : Jean-Marie Pierrel et Laurent Romary. Jean-Marie Pierrel, qui m'a fait confiance et permis de réaliser cette thèse, dans les meilleures conditions, au sein de l'équipe Dialogue qu'il anime.

Laurent Romary, dont les qualités humaines et scientifiques ne sont plus à démontrer. Il a su amorcer cette réflexion, la guider jusqu'à sa maturité. Je suis convaincu que le fait d'avoir travaillé sous sa direction, aura été pour moi une très grande chance.

Je tiens à remercier Yvon Gardan sans qui ce manuscrit n'aurait sans doute jamais vu le jour. Il m'a permis de m'engager dans la voie de la recherche en informatique.

Ma gratitude va également aux personnes qui ont contribué à l'amélioration de la qualité tant syntaxique que sémantique de ce mémoire : Bertrand Gaiffe pour ses lectures « temps réel » et ses critiques pointues ; Yannick Lallement, pour ses lectures « outre-atlantiques », il a su donner à ce mémoire une lisibilité, que seul je n'aurais pu lui procurer, enfin, merci à Jean-Christophe Dubois pour ses lectures finales.

Tous mes remerciements vont à ceux qui ont contribué à rendre cette période agréable et sympathique : Nadia Bellalem, Lionel Beaugé, Emmanuel Bouyer, Laurent Chapelier, Céline Cordier, Stéphane Durand, Jean-Michel Gallone, Norbert Glaser, Anne Reboul, Gilles Souvay et Antoine Tabbone. Merci également à Chantal, pour sa gentillesse.

Merci à mes collègues enseignants pour leur soutien et leur aide : Jeanine Souquières, Anne Boyer, Azim Roussanaly et Christelle Scharff.

Merci également à mes amis, proches ou lointains, déjà cités ou non, pour leur présence, ainsi qu'à ma famille et à mes parents, dont le soutien sans faille tout au long de ces années m'a été plus que précieux : indispensable.

Enfin, merci à toi Stéphanie, pour ton soutien, ton aide et tes encouragements, notamment durant les périodes difficiles. Sans toi, rien n'aurait été pareil...

*À tous ceux qui me sont chers,
à Stéphanie.*

*«L'imagination est plus importante que la
connaissance» (A. Einstein).*

Table des matières

Résumé	1
Remerciements	3
Préambule	13
Introduction	15

1

Trois fils conducteurs

1.1	Introduction	20
1.2	Perspective de dialogue	20
1.2.1	La notion de dialogue en psychologie	20
1.2.2	La notion de dialogue à travers la linguistique et la psycholinguistique	21
1.2.3	La notion de dialogue en ergonomie	23
1.2.4	La notion de dialogue en informatique	24
1.2.5	Conclusion	25
1.3	Le contexte	25
1.3.1	Le contexte en psychologie sociale	26
1.3.2	Le contexte en pragmatique	27
1.3.3	Le contexte en dialogue homme-machine	29
1.3.4	Conclusion	30
1.4	Principe d'économie cognitive	31
1.4.1	Le principe d'économie cognitive à la base de la cognition humaine?	32
1.4.2	Le principe d'économie cognitive en pragmatique	34
1.4.3	Application au niveau du dialogue homme-machine	35
1.4.4	Réflexion sur l'application du principe d'économie cognitive	39

1.4.5	Conclusion	40
1.5	Illustration du tryptique dialogue, contexte et principe d'économie cognitive	41
1.6	Synthèse	43

2

Modélisation des usages spatiaux de la langue
--

2.1	Introduction	46
2.2	Limites d'une approche géométrique	47
2.2.1	L'exemple de " <i>près de/loin de</i> "	47
2.2.2	L'exemple de la préposition " <i>dans</i> "	48
2.2.3	Conclusion	49
2.3	La prise en compte des aspects fonctionnels: une nécessité	49
2.3.1	La compositionnalité	49
2.3.2	La transitivité	50
2.3.3	Conclusion	50
2.4	Le modèle de Vandeloise	50
2.4.1	Prolégomènes: les notions de ressemblance de famille et d'impulsion	51
2.4.2	La prise en compte de la pragmatique chez Vandeloise	52
2.4.3	Les aspects fonctionnels	54
2.4.4	Un exemple: les prépositions " <i>sur/sous</i> "	56
2.4.5	Conclusion	58
2.5	Le modèle de Asher et Sablayrolles	59
2.5.1	Introduction	59
2.5.2	Taxinomie des verbes de mouvement	59
2.5.3	Sémantique des verbes de changement de lieu	61
2.5.4	Conclusion	64
2.6	Le modèle de Hernández	64
2.6.1	Les bases du modèle de Hernández	65
2.6.2	Les relations topologiques	65
2.6.3	Les relations d'orientation	66
2.6.4	Prise en compte des objets possédant une extension spatiale	67
2.6.5	Illustration	68
2.6.6	Conclusion	69
2.7	Synthèse	70

3

Le problème de la référence

3.1	Introduction	72
3.2	Le problème de la référence en linguistique	73
3.3	Le traitement de la référence en informatique	74
3.3.1	Prolégomènes	74
3.3.2	Le problème de la référence aux actions	76
3.4	Le problème de la référence aux objets en informatique	78
3.4.1	La DRT de Kamp	78
3.4.2	Le focus de Sidner	81
3.4.3	Un modèle de calcul des référents fondé sur la notion d'axiologie	82
3.5	Problématique et intérêt de la référence spatiale	88
3.5.1	Pourquoi s'intéresser aux références spatiales purement langagières?	89
3.5.2	Notre contribution	91
3.5.3	Conclusion	97
3.6	Une étude détaillée de la préposition "à gauche de"	98
3.6.1	Un modèle discrétisant l'espace	98
3.6.2	Un modèle linguistique	99
3.6.3	Un modèle à base de nuages d'acceptabilité	100
3.6.4	Un modèle fondé sur des inclusions de portions de surface	102
3.6.5	Un modèle à base de zones d'acceptabilité	103
3.6.6	Le modèle de Hernández	103
3.6.7	Autres limites de ces modèles	103
3.6.8	Conclusion	105
3.7	Vers une autre approche	105
3.7.1	Introduction	105
3.7.2	Rôle possible de la notion de saillance	106
3.7.3	Rôle possible de la notion de prototype	106
3.7.4	Vers un modèle de calcul des référents spatiaux	109
3.8	De la nécessité d'introduire la notion de cadre	121
3.9	Conclusion	122

4

Vers un nouveau concept, celui de cadre

4.1	Introduction	126
4.2	De la nécessité des cadres	127
4.2.1	Une nécessité au niveau spatial	127
4.2.2	Une nécessité au niveau informatique	128
4.2.3	Conclusion	130
4.3	Définition et caractérisation de la notion de cadre	131
4.3.1	Définition générale	131
4.3.2	Les cadres et le problème de la référence spatiale	132
4.3.3	Tout est “ <i>alternative</i> ” !	133
4.3.4	Caractérisation des cadres de référence	135
4.3.5	Évolution discursive des cadres de référence	138
4.4	Les cadres de référence et l’existant	140
4.4.1	Les cadres de référence et l’approche de Vandeloise	140
4.4.2	Les cadres de référence et l’approche de Asher et Sablayrolles	142
4.4.3	Les cadres de référence et les aires de recherche de Pribbenow	143
4.4.4	Les cadres de référence et les univers spatiaux de Charolles	144
4.4.5	Conclusion	146
4.5	Connaissances nécessaires à l’interprétation d’une référence spatiale	147
4.5.1	Le nouveau cadre de référence	147
4.5.2	L’historique du dialogue et le lexique	159
4.5.3	Analyse syntaxico-sémantique	162
4.5.4	La perception	173
4.5.5	L’application	178
4.6	Principe de fonctionnement de la notion de cadre de référence	179
4.7	Application de la notion de cadre de référence à un système de représentation graphique	183
4.7.1	Introduction	183
4.7.2	Exemple détaillé	183
4.7.3	Le modèle de la notion de cadre de référence: un bilan	191
4.8	Prémises pour un modèle de traitement des énoncés de positionnement	192
4.8.1	Introduction	192
4.8.2	Vers un positionnement plus “cognitif”, le cas de “ <i>ici</i> ”	193
4.8.3	Quelques pistes pour formaliser les énoncés de positionnement	195
4.8.4	Conclusion	197

4.9	Autres validations des cadres de référence	198
4.9.1	Modèles informatiques, modèles de la cognition humaine: un cas de divorce?	198
4.9.2	Validation psychologique des cadres de référence	199
4.9.3	Cadres de référence et ambiguïté	200
4.9.4	Les cadres de référence et la transitivité	202
4.9.5	Liens avec le temps	202
4.10	Conclusion	203

5

Intégration d'une composante spatiale dans un système de dialogue

5.1	Introduction	205
5.2	Les systèmes de dialogue	206
5.2.1	Fonctions à mettre en œuvre dans un système de dialogue	206
5.2.2	Les connaissances à mettre en œuvre dans un système de dialogue	208
5.2.3	Bilan	209
5.3	Intégration d'une composante spatiale dans un système de dialogue	209
5.3.1	“Un fondement solide” : le modèle de CAO	210
5.3.2	Bilan de ce modèle	213
5.4	La maquette MultiDial	215
5.5	Conclusion	222

Conclusion et perspectives

Annexes

A	L'expérience de Milgram	229
B	La pensée latérale	233
C	Élaboration des groupes homogènes perceptifs	235
C.1	Construction des groupes homogènes perceptifs de type <i>droite/gauche</i>	235
C.2	Construction des groupes homogènes perceptifs de type <i>haut/bas</i>	238
C.3	Construction des groupes homogènes perceptifs de proximité	238
	Table des figures	239

Bibliographie

243

Glossaire

253

Préambule

Les phénomènes que nous étudierons dans cette thèse sont majoritairement d'ordre linguistique, ce qui ne saurait surprendre si l'on admet la prépondérance du langage naturel en tant que mode de communication dans un dialogue homme-machine. Pourtant, il n'est pas difficile de comprendre que le chercheur se trouve ici dans la délicate position de mener à bien la mise en œuvre de tels systèmes, qui reposent sur une bonne compréhension préalable de la langue et de ses mécanismes, sans pour autant faire œuvre de linguiste. Ainsi, bien que nous autres, informaticiens travaillant sur le langage, soyons attentifs aux mêmes phénomènes que le linguiste, notre visée est différente : ce n'est pas comprendre la langue en tant que telle mais produire une représentation de ses mécanismes, qui puisse parfois, à moyen terme, conduire à la réalisation de systèmes dans de bonnes conditions. Mais inversement, le travail théorique qui précède l'implémentation sur machine ne doit pas être négligé. Les modèles que nous proposons ne doivent en aucun cas entrer en contradiction avec la réalité de la langue, et le linguiste est souvent là pour nous le rappeler.

Note : le fait d'avoir essayé, quand cela a été possible, de donner un éclairage psychologique à nos travaux nous incite à reprendre dans une certaine mesure les remarques qui précèdent pour les appliquer au domaine spécifique de la psychologie.

Introduction

Le travail présenté dans cette thèse porte sur un ensemble de réflexions liées à la définition de systèmes de dialogue homme-machine. Jusqu'à présent, l'informatique nous a habitués à des échanges qui s'appuyaient essentiellement sur une ou deux modalités d'interaction, notamment à partir de textes affichés ou tapés au clavier et d'icônes pointées à l'aide de la souris. L'inconvénient de ce type d'interaction provient des nombreuses contraintes qu'elle impose à l'utilisateur. Aussi, assiste-t-on, depuis quelques années, à une inversion de tendance : après une phase où les développements informatiques ont été guidés par les avancées scientifiques et technologiques et ont conduit à des systèmes auxquels l'homme a dû s'adapter, il convient aujourd'hui de redonner la priorité à l'homme en adaptant les systèmes informatiques aux besoins et potentialités humains¹. Dès lors, si l'on se fixe comme objectif de minimiser l'effort d'adaptation de l'utilisateur à la machine, le choix de la langue naturelle semble s'imposer. Celle-ci présente en effet l'avantage d'être une compétence déjà acquise par l'utilisateur et un moyen de communication particulièrement puissant surtout lorsqu'il s'agit d'exprimer, pour la réalisation d'une tâche^(*)², un but complexe qu'on ne peut traduire directement en termes d'opérateurs de l'application.

Cette thèse présente précisément un ensemble de recherches qui visent à améliorer l'interaction entre l'homme et la machine, nos efforts ayant porté plus spécifiquement sur les aspects spatiaux, notamment sur le problème de la référence spatiale en langage naturel. Avant de présenter rapidement le contenu de ce manuscrit, il nous semble important de préciser la démarche qui a guidé notre travail.

La réflexion menée dans cette thèse est née d'un besoin qui est apparu en particulier au sein de la maquette MultiDial, issue du projet Esprit MultiWorks qui a été développée dans notre équipe. Cette maquette a pour rôle d'assurer la gestion du dialogue entre un utilisateur et une application informatique complexe, le pilotage de l'application étant assuré par une entrée vocale (cette entrée vocale étant simulée pour l'instant au moyen du clavier). La maquette MultiDial a alors été testée en grandeur nature afin de piloter le logiciel MultiCard³. Ces tests ont mis en lumière certaines lacunes du système MultiDial, en particulier, son inaptitude à prendre en compte les requêtes spatiales formulées par l'utilisateur. L'adjonction à la maquette MultiDial d'un module capable d'interpréter les références spatiales s'avérait donc nécessaire afin d'augmenter la souplesse du dialogue. Au sein de MultiCard, la représentation graphique des objets manipulés, étant relativement

1. On dénomme ce type d'approche la « conception centrée sur l'utilisateur ».

2. Les mots marqués d'un astérisque : (*) sont définis dans le glossaire (page 253).

3. MultiCard est un logiciel qui permet de manipuler des documents texte, audio, vidéo et de naviguer au sein de tels documents.

similaire, il nous a paru opportun d'élargir notre réflexion au problème de calcul des référents spatiaux dans un système de visualisation en deux dimensions. Mais, là encore, le caractère fortement contraint de la tâche nous a incité à reconsidérer le problème de la modélisation de l'espace en langage naturel dans sa totalité pour nous recentrer ensuite sur celui de la référence spatiale tout en restant à un niveau de généralité suffisant. Par ailleurs, lors de nos recherches, nous avons été attentifs à l'émergence d'éventuels paradigmes plus généraux sous-jacents à la réflexion menée dans cette thèse. Nous avons pu dégager ainsi trois grands thèmes : la notion de dialogue, de contexte et enfin celle de principe d'économie cognitive.

Notre étude débutera donc par une illustration de l'importance des notions de dialogue, de contexte et de principe d'économie cognitive (*chapitre 1*). Ces trois concepts reviendront tout au long de cette thèse et joueront le rôle de fils conducteurs. Qui plus est, nous tenterons de montrer que ce tryptique émerge sensiblement de la linguistique, de l'informatique et de la psychologie sociale.

Dans un second temps, nous chercherons les premiers points clés sur lesquels doit reposer une modélisation de l'espace en langage naturel (*chapitre 2*). Cette recherche nous amènera à présenter trois modèles qui ont influé sur notre réflexion. Ces modèles qui relèvent respectivement des approches linguistiques, logico-computationnelles et computationnelles nous permettront également d'étudier certains aspects sur lesquels la communauté, qui s'intéresse à la modélisation de l'espace en langage naturel, s'appuie de plus en plus. Il s'agit en particulier de l'importance de la conception d'un espace relatif et non absolu, qui repose pour une bonne part sur la notion de cible^(*) et de site^(*). Par ailleurs, nous insisterons sur le rôle croissant des aspects fonctionnels au sein des modélisations spatiales.

La partie suivante nous permettra d'étudier la problématique de la référence en langage naturel. Ainsi, étant donné que le domaine de la référence spatiale relève d'une sous-problématique spécifique à celle de la référence en langage naturel, nous en viendrons tout naturellement à présenter cette dernière ainsi que les travaux qui y ont été menés (*chapitre 3*). Nous nous recentrerons ensuite sur des aspects plus spatiaux pour nous intéresser au pilotage d'un système de visualisation graphique en deux dimensions. Nous insisterons alors sur le fait qu'une bonne interface homme-machine dédiée à la manipulation d'un univers graphique doit permettre à l'utilisateur de formuler trois types de requêtes : les requêtes purement gestuelles, les requêtes combinant geste et parole et enfin les requêtes orales. La suite de cette thèse s'inscrira dans le cadre de ces dernières et plus particulièrement de la compréhension des requêtes orales purement spatiales (l'entrée vocale étant simulée par le clavier). Nous démontrerons en particulier que de telles requêtes trouvent parfaitement leur place au sein d'une interaction homme-machine par rapport à des requêtes combinant le geste et la parole.

Ces préoccupations nous amèneront à rechercher au sein de la littérature une approche qui modélise les prépositions spatiales projectives⁴. Pour simple que puisse paraître l'uti-

4. On distingue les prépositions *projectives* ("à droite", "devant", "dessus"...), les prépositions *topolo-*

lisation de ces prépositions, nous verrons que les modèles actuels rencontrent bien des difficultés. Ce constat nous amènera à esquisser un nouveau modèle, la présentation de ce dernier sera précédée d'une réflexion portant sur l'applicabilité des notions de saillance et de prototype.

Au vu des nombreuses restrictions apportées au modèle proposé dans le chapitre 3, il nous est apparu nécessaire d'élargir notre réflexion à la modélisation générale de l'espace en langage naturel (*chapitre 4*). Après avoir défini la notion de *cadre* qui semble suffisamment générale pour s'appliquer au sein de tout type de raisonnement spatial, nous en viendrons à proposer la notion de *cadre de référence* qui s'applique au domaine spécifique de la référence spatiale en langage naturel. Nous poserons ensuite les principaux jalons de cette notion qui se décline suivant des spécificités perceptives, fonctionnelles et sémantiques. Ces spécificités nous amèneront à décrire précisément sur quelles connaissances se fonde l'élaboration des cadres de référence. Nous illustrerons leur fonctionnement pour le cas spécifique du pilotage d'un système graphique et nous achèverons ce chapitre par une discussion portant sur l'applicabilité de la notion de cadre de référence à la prise en compte des énoncés de positionnement.

La dernière partie de ce mémoire traite de l'intégration de la notion de cadre de référence au sein d'un système de dialogue homme-machine. Après une présentation des spécificités générales sous-jacentes à l'élaboration d'un système de dialogue homme-machine, nous décrirons la réalisation d'une plateforme graphique qui s'inspire d'aspects relatifs à la conception assistée par ordinateur. Nous décrirons également la maquette MultiDial au sein de laquelle seront intégrées la plate-forme graphique ainsi que la méthodologie de recherche des référents spatiaux sur la base de la notion de cadre de référence.

giques (“à”, “sur”, “contre”, “dans”, etc), les prépositions *géométriques* (“entre”, “parmi”) et les prépositions dynamiques (“de”, “à”, etc).

Chapitre 1

Trois fils conducteurs

Sommaire

1.1	Introduction	20
1.2	Perspective de dialogue	20
1.2.1	La notion de dialogue en psychologie	20
1.2.2	La notion de dialogue à travers la linguistique et la psycholinguistique	21
1.2.3	La notion de dialogue en ergonomie	23
1.2.4	La notion de dialogue en informatique	24
1.2.5	Conclusion	25
1.3	Le contexte	25
1.3.1	Le contexte en psychologie sociale	26
1.3.2	Le contexte en pragmatique	27
1.3.3	Le contexte en dialogue homme-machine	29
1.3.4	Conclusion	30
1.4	Principe d'économie cognitive	31
1.4.1	Le principe d'économie cognitive à la base de la cognition humaine?	32
1.4.2	Le principe d'économie cognitive en pragmatique	34
1.4.3	Application au niveau du dialogue homme-machine	35
1.4.4	Réflexion sur l'application du principe d'économie cognitive	39
1.4.5	Conclusion	40
1.5	Illustration du tryptique dialogue, contexte et principe d'économie cognitive	41
1.6	Synthèse	43

1.1 Introduction

Si l'on se pose la question de savoir ce que d'autres disciplines telles que la linguistique ou la psychologie pourraient apporter à l'informatique et plus spécifiquement au dialogue homme-machine, il semble que l'on puisse dégager trois thèmes récurrents. Comme nous le verrons au cours de ce chapitre, ces thèmes émergent significativement de la linguistique, de la psychologie et de l'informatique et sous-tendent la réflexion qui est exposée dans ce mémoire.

Ainsi, outre la notion de cadre que nous développerons tout au long de ce mémoire, trois fils conducteurs sont sous-jacents à cette thèse :

- le fait que nous nous plaçons dans une perspective de dialogue,
- la notion de contexte,
- le principe d'économie cognitive.

Nous tenterons de démontrer le rôle essentiel que jouent ces trois thèmes dans le domaine de la psychologie, celui de la pragmatique et enfin celui de l'informatique (la notion de dialogue sera également illustrée dans une perspective ergonomique). Ce chapitre ne constitue cependant pas une taxinomie complète de toutes leurs applications possibles, il vise surtout à démontrer leur importance croissante à l'aide de quelques éclairages ponctuels, tout en illustrant chaque fois que possible leur influence sur nos travaux. Le fil conducteur de ce chapitre pourrait donc s'intituler "Qu'est-ce que d'autres disciplines concomitantes à l'informatique pourraient apporter au dialogue homme-machine?".

Nous terminerons ce chapitre en relatant les travaux de Garrod [1987] et [1994] qui nous fourniront l'occasion de donner un éclairage unifié des trois notions abordées.

1.2 Perspective de dialogue

Cette partie se propose de fournir plusieurs éclairages sur la notion de dialogue et ce, au sein du champ pluridisciplinaire des sciences cognitives.

1.2.1 La notion de dialogue en psychologie

Comme le souligne Vivier [1997], il y a cinq ou dix ans, l'idée ne serait venue à personne de s'attarder sur des problèmes psychologiques à propos de dialogue homme-machine. La majorité des études dont nous disposons à l'heure actuelle relève des problèmes que rencontrent les utilisateurs pour s'adapter à l'ordinateur mais très peu de travaux s'intéressent au fonctionnement du partenaire humain, de son langage, de ses conduites et de sa réflexivité¹. Il est vrai que jusqu'à présent, il a suffi que les procédures inventées par l'informatique permettent d'obtenir de la machine l'efficacité visée en se moquant (ou presque) des moyens utilisés.

J. Vivier souligne par ailleurs que le dialogue homme-machine relève à certains égards plutôt d'une communication homme/homme utilisant la médiation de la machine

1. Réflexivité : faculté pour un être humain de représenter son propre comportement et de raisonner à son propos.

dans le sens où les deux auteurs de cette communication sont le programmeur (ou plus précisément les équipes de concepteurs ayant réalisé les différentes couches de logiciels) et l'utilisateur. De notre point de vue, cet aspect est de moins en moins vrai dans le sens où les machines sont de plus en plus souvent dotées de moyens d'apprentissage qui leur confèrent la possibilité de venir à bout de certains problèmes qui n'ont pas été prévus explicitement par le concepteur.

Nous nous sommes posé la question de savoir si une voie prometteuse à plus long terme ne pourrait pas consister à s'intéresser aux très nombreux travaux qui ont été effectués dans le domaine de la communication humaine qui vise à l'étude du dialogue homme-homme. On pourrait voir alors le dialogue homme-machine comme une spécialisation du dialogue homme-homme.

On trouve en effet au sein des travaux portant sur la communication nombre de modèles qui vont même jusqu'à expliquer certains comportements et sentiments humains comme la PNL (programmation neurolinguistique) [Cudicio, 1986] ou l'analyse transactionnelle [Berne, 1990], [Woollams *et al.*, 1990] et [Marc, 1993].

Face à la richesse de tels modèles eut égard à ce dont nous autres informaticiens avons besoin pour étudier un "simple" dialogue homme-machine, on peut s'interroger sur la possibilité de recadrer ces modèles au niveau du dialogue homme-machine.

Nous pensons que ce recadrage se heurte pour l'instant à la finalité toute différente des études en communication en regard de l'étude dont nous aurions besoin pour le dialogue homme-machine. En effet, ces études de la communication ne visent qu'à comprendre et décrire les phénomènes et non pas à les formaliser. On peut faire ici un parallèle entre le couple communication/informatique et le couple linguistique/informatique. Ainsi, la linguistique, de par sa méthodologie centrée essentiellement sur l'analyse de l'acceptabilité ou non d'énoncés attestés ou construits, n'a pas porté une attention particulière aux circonstances d'élocution de tels énoncés, or c'est précisément ce type de formalisation dont les informaticiens souhaiteraient disposer.

Comme J. Vivier, nous concluons cette partie en remarquant que le dialogue homme-machine ne manque pas de questions pour les psychologues, ne serait-ce que parce que un humain y intervient. Voyons à présent un éclairage linguistique de la notion générale de dialogue.

1.2.2 La notion de dialogue à travers la linguistique et la psycholinguistique

Ce paragraphe va nous fournir l'occasion de glisser de la linguistique à la psycholinguistique en passant des travaux d'Austin à ceux de Searle. Pour ce faire, il nous faut tout d'abord situer ces deux disciplines. Ainsi, en psycholinguistique, la méthodologie est celle d'une science expérimentale dans le sens où on soumet un certain nombre de sujets à des tests, qui ont pour fonction de mettre à l'épreuve des hypothèses formulées indépendamment. En revanche, en linguistique, la matière première ne provient pas toujours d'un corpus. Le linguiste produit alors lui-même les données dont il a besoin et les décrit à l'intérieur d'un cadre théorique autonome et complet.

À la base des études psycholinguistiques pouvant fournir une contribution à l'étude du dialogue homme-homme et plus particulièrement du dialogue homme-machine, se trouve la théorie des actes de langage développée par Austin [1970]. Austin, en remettant en cause le postulat de *l'illusion descriptive*², défendait l'idée selon laquelle toute phrase grammaticale complète correspond à la performance de trois actes différents :

- un *acte locutoire*, qui correspond à la prononciation par le locuteur de telle ou telle phrase.
- un *acte illocutoire*, qui correspond à l'acte que le locuteur entendait accomplir en utilisant telle ou telle phrase. On dénombre cinq classes d'actes illocutoires : les verdictifs, les exercitifs, les promissifs, les comportatifs et les expositifs (voir [Moeschler et Reboul, 1994] pour un exposé détaillé).
- un ou plusieurs *actes perlocutoire(s)*, qui sont des actes que le locuteur accomplit par la prononciation de telle ou telle phrase : la persuasion, l'intimidation, etc.

Searle [1972] reprend cette théorie et propose de distinguer deux éléments de la structure syntaxique de la phrase : le marqueur de *contenu propositionnel* et le marqueur de *force illocutoire*, le premier indique la proposition exprimée, alors que le second indique l'acte illocutoire accompli. Ainsi, prenons l'exemple des deux énoncés suivants :

Jean fume beaucoup. (1)
Jean fume-t-il beaucoup? (2)

Si c'est la même proposition qui est exprimée en (1) et (2), chacun de ces énoncés accomplit un acte illocutoire différent, soit respectivement une affirmation et une question. Searle propose la notation $\mathcal{F}(\mathcal{P})$ où \mathcal{F} correspond au marqueur de force illocutoire et \mathcal{P} à la proposition. Searle aboutit finalement à sa propre classification des actes illocutoires : les représentatifs, les directifs, les promissifs, les expressifs et les déclaratifs (voir [Moeschler et Reboul, 1994] pour un exposé détaillé).

Nous n'en dirons pas plus sur les fondements de ces travaux sinon que leur influence est de plus en plus sensible dans la réalisation des interfaces homme-machine, en autorisant en particulier une prise en compte plus fine des intentions du locuteur. On peut citer à ce propos les travaux réalisés dans notre équipe par Chapelier [Chapelier *et al.*, 1994], [Chapelier *et al.*, 1995] et [Chapelier, 1996] sur l'étude des différents besoins d'assistance dont un utilisateur peut souhaiter disposer dans la manipulation d'une interface informatique très complexe, en l'occurrence le logiciel GOCAD [Mallet, 1992]. Ce dernier autorise la modélisation et la manipulation de surfaces naturelles géologiques.

Chapelier n'a recours pour son modèle d'analyse qu'aux trois forces illocutoires suivantes de la théorie de Searle :

- la force *représentative* : son but est d'engager le locuteur à des degrés divers sur la vérité de la proposition exprimée.

2. *Illusion descriptive* : la fonction des phrases déclaratives ne se limite pas seulement à décrire le monde, c'est-à-dire qu'elles ne sont pas soit vraies, soit fausses.

- la force *promissive* : son but est d'engager le locuteur à accomplir une action future.
- la force *directive* : son but est d'obtenir que l'auditeur fasse quelque chose. Concernant cette dernière force, Chapelier distingue l'ordre (O) et la requête (R). Dans un contexte d'interaction homme-machine, la première force illocutoire, contrairement à la seconde, nécessite pour sa satisfaction une modification de l'état du monde qui se trouve affiché à l'écran (dans le contexte spécifique du logiciel GOCAD).

Voici une illustration de ces différentes forces sur un extrait du corpus que Chapelier a recueilli à l'aide du paradigme de type magicien d'Oz³:

Locuteur	Énoncé	Force illocutoire
<i>Utilisateur :</i>	<i>Faire une interpolation</i>	Directive O
Magicien d'Oz :	Sur quelle surface?	Directive R
<i>Utilisateur :</i>	<i>Sur la surface A</i>	Représentative
Magicien d'Oz :	J'interpole la surface	Promissive
Magicien d'Oz :	Surface interpolée	Représentative

La théorie des actes de langages doublée d'une méthode d'analyse conversationnelle, ont permis à Chapelier d'identifier clairement les différentes forces d'assistance apparues dans le corpus qu'il a recueilli. Ces différentes formes qui s'actualisaient au cours du dialogue ont été formalisées au sein d'un modèle de dialogue d'assistance générique implémentable dans un système informatique.

Nous retiendrons donc de la théorie des actes de langage son rôle de guide dans la recherche de l'intention du locuteur.

1.2.3 La notion de dialogue en ergonomie

L'ergonomie réalise l'étude scientifique des conditions (psychophysiologiques et socio-économiques) de travail et des relations entre l'homme et la machine en cherchant à les améliorer. Ainsi, tandis que le dialogue homme-homme est le terrain privilégié du psycholinguiste, celui du dialogue homme-machine est celui de l'ergonome.

L'outil de prédilection des ergonomes pour étudier le comportement d'un sujet humain en situation de communication avec un ordinateur n'est autre que le paradigme du magicien d'Oz dont nous venons de voir une illustration à l'aide des travaux de Chapelier. La question que se sont posée les ergonomes a été de savoir si le dialogue homme-machine diffère de celui du dialogue homme-homme, et si oui en quoi?

Leurs conclusions aboutissent à des différences significatives⁴ entre le dialogue homme-machine et le dialogue homme-homme dont nous rappelons ici les principales d'après

3. Paradigme du magicien d'Oz : expérience consistant à recueillir des dialogues dans lesquels la machine qui, à terme, sera l'un des interlocuteurs du dialogue, est simulée par une ou plusieurs personne(s) appelée(s) classiquement le(s) compère(s).

4. Il convient toutefois de se méfier de certaines expériences dont on a parfois pu déduire tout et n'importe quoi. Les erreurs qui ont pu être faites proviennent bien souvent de la négligence de certains aspects qui ont pu influencer de manière déterminante l'étude qui a été faite de l'interaction homme-machine, voir en particulier Vilnat [1997] pour une discussion portant sur ces problèmes.

Amalberti [Amalberti *et al.*, 1993] et Nicolle [1997] :

- la complexité des phrases est appauvrie,
- le langage devient normé et ressemble à de l'écrit,
- en cas de problème de communication, le sujet n'effectue généralement pas de reformulation et préfère une répétition mécanique.

De notre point de vue, ces différences ont de quoi surprendre si l'on se réfère à la fameuse expérience menée avec le logiciel ELIZA. ELIZA est un programme qui peut être écrit en Basic en quelques heures et qui donne à un utilisateur novice l'illusion d'un véritable dialogue avec un psychiatre. Voici un extrait d'un dialogue type :

La machine: *Bonjour, comment allez-vous ?*

Utilisateur: *Je ne me sens pas très bien aujourd'hui.*

La machine: *Et les autres jours ?*

Utilisateur: *Ça dépend, mais sauriez-vous diagnostiquer mon mal ?*

La machine: *Me croyez-vous capable de diagnostiquer votre mal ? ...*

Le secret de ce système réside dans un filtrage par mots clés avec un renvoi d'un écho quelque peu modifié. Si le système ne sait que répondre, il s'en sort en énonçant une phrase passe-partout comme par exemple : "*dites m'en plus*".

Durant cette expérience, les utilisateurs et même certains psychiatres ont eu véritablement l'impression de dialoguer, au moyen du clavier, avec une machine qui avait la capacité de les aider psychologiquement. Il va de soi que cette capacité d'aide psychologique dépasse de loin les performances nécessaires à l'établissement d'un simple dialogue de type commande de processus. Il nous semble alors étonnant qu'un utilisateur puisse prêter des capacités aussi évoluées à une machine (en lui fournissant des phrases très riches sémantiquement^(*)) alors qu'il se contente de lui soumettre des phrases très pauvres syntaxiquement⁵ (de peur sans doute que la machine ne comprenne pas!?).

En prenant un peu de recul, on peut remarquer que cette expérience illustre un principe très récurrent actuellement dans la littérature du dialogue homme-machine, à savoir l'importance de *comment* l'homme perçoit la machine. La réponse à cette question va conditionner pour beaucoup les capacités et facultés que l'homme octroie à l'ordinateur et influencer alors directement sur le dialogue qu'il aura avec ce dernier. Cette remarque nous conduit à parler de la notion de dialogue en informatique.

1.2.4 La notion de dialogue en informatique

La notion de dialogue en informatique pose la question suivante : pourquoi faire communiquer l'homme et la machine au moyen du langage plutôt qu'à l'aide d'une interface faite de boutons et de menus? L'ouvrage de Nicolle *et al.* [1997] apporte quelques éléments de réponse à cette question : la langue est pour l'être humain l'outil de communication dont il a la plus grande maîtrise, cette dernière résulte d'années d'apprentissage que ne

5. Les travaux de Amalberti [1993] démontrent que la restriction au niveau de la longueur des phrases saisies ne peut s'expliquer seulement par le fait que ces dernières étaient saisies au moyen du clavier. Qui plus est, le logiciel Eliza posait des questions ouvertes (par opposition aux questions fermées auxquelles on répond par "*oui*" ou par "*non*").

remplaceront jamais un manuel d'instructions relatif à une interface conçue de façon artificielle, aussi élaborées que puissent être les fonctionnalités offertes. Par ailleurs, une interface homme-machine basée sur le langage profiterait également de l'adéquation des langues naturelles au fonctionnement de l'esprit humain [Pierrel et Romary, 1997]. On peut remarquer que ces deux principes s'inscrivent dans l'aspiration principale du dialogue homme-machine qui va vers une plus grande adaptation de la machine à l'homme et non l'inverse.

L'utilité d'introduire la notion de dialogue apparaît également dans la distinction qui est faite entre le "faire" et le "faire faire" que Pouteau [1995] (entre autres auteurs) met notamment en avant. Si le "faire" peut se contenter d'une modalité gestuelle, en revanche, le "faire faire" semble plus en adéquation avec une modalité à forte dominante orale. Par ailleurs, étant donné que jusqu'à présent, on ne peut demander à l'ordinateur d'accomplir des requêtes trop complexes, le "faire faire" se place plus dans une perspective de coopération entre l'homme et la machine en vue de réaliser une tâche donnée. La notion de "faire faire" opposée au "faire" ainsi que la nécessité d'une coopération entre l'homme et la machine nous placent donc réellement dans une perspective de dialogue.

1.2.5 Conclusion

De ce paragraphe, nous retiendrons tout d'abord le caractère pluridisciplinaire de la notion de dialogue. Du côté spécifique du dialogue homme-machine, il faudra sans doute retenir que ce dernier offre des singularités qui en font autre chose qu'un dialogue humain. Ce fait prêche en faveur d'une étude encore plus approfondie de la part des linguistes et des psychologues du dialogue homme-machine. Vivier [1997] souligne d'ailleurs que les psychologues ont pris un certain retard par rapport aux linguistes dans ce domaine.

Nous retiendrons également l'importance déterminante de l'intention du locuteur, que tente de cerner la théorie des actes de langage. Par ailleurs, du côté de l'utilisateur, la vision des capacités qu'il accorde à sa machine aura une influence déterminante sur la manière dont il réalisera le(s) but(s) qu'il s'est fixé(s).

Enfin, pour ce qui nous concerne, nous avons souligné la nécessité d'un dialogue homme-machine qui soit le plus naturel possible. Dans ce sens, nous chercherons au cours du chapitre 3 à autoriser plus de liberté à l'utilisateur, dans sa formulation de requêtes spatiales purement langagières portant sur un système d'affichage graphique.

1.3 Le contexte

La notion de contexte recouvre l'ensemble des circonstances dans lesquelles s'insère un fait. Cette notion se situe au carrefour des sciences cognitives, comme cela a été mis en évidence lors du colloque interdisciplinaire de Cognisciences-Est à Strasbourg [Pierrel, 1992]. Les disciplines comme la psychologie, la pragmatique et l'informatique y font largement appel afin d'expliquer régulièrement de nouveaux phénomènes.

Les difficultés inhérentes à la notion de contexte proviennent sans doute de son caractère trop général qui ne permet pas de décrire les phénomènes de manière suffisamment fine. C'est pourquoi nous étudierons cette notion sous l'angle spécifique de la psychologie sociale qui est connexe au dialogue homme-machine, celui de la pragmatique et enfin celui de l'informatique.

1.3.1 Le contexte en psychologie sociale

Les psychologues ont récemment redécouvert l'importance déterminante de la situation dans laquelle se trouve placé un patient donné au travers notamment des travaux de Leyens [1992]. Celui-ci démontre que la minimisation de l'importance du rôle joué par la situation peut conduire à bien des erreurs.

Il introduit ce qu'il appelle "l'erreur fondamentale des psychologues", cette erreur que nous faisons tous et que commettent certains psychologues est cette tendance, véritablement ancrée en nous, qui fait que nous avons une propension à sur-estimer, dans nos explications, la part qui provient de l'individu, les causes internes, la personnalité et à sous-estimer celle qui résulte de la situation, des causes externes, des circonstances. La célèbre expérience de Milgram⁶ illustre très bien ce propos (les principes et résultats de cette surprenante expérience sont exposés en détail à l'annexe A). Milgram, a conclu de cette expérience que nous sommes tous de potentiels "monsieur ou madame 450 volts", non pas à cause de notre personnalité ou parce que nous avons répondu d'une certaine façon à un test, mais parce que, mis dans certaines situations, il y a de grandes chances pour que celles-ci nous entraînent à agir de manière déterminée.

De notre côté, nous ferons un parallèle très net entre la notion de situation décrite par Leyens et celle de contexte. Ce dernier point semble démontrer que le contexte a un rôle très important à jouer au sein de la psychologie sociale. Dès lors, on peut se demander si cette illustration du rôle du contexte au sein de la psychologie sociale est un cas isolé ou non. L'étude de "l'effet Pygmalion" tel qu'il a été mis en évidence par Rosenthal et Frode [1963] semble pouvoir apporter quelques éléments de réponse.

Nous expliciterons cet effet à l'aide de l'une des premières expériences que Rosenthal a conduite sur des rats blancs. Dans le cadre d'un cours de psychologie, des étudiants doivent s'exercer à la manipulation d'animaux. Chacun d'eux aura la responsabilité d'un seul rat qu'il devra familiariser avec un labyrinthe, remettre en début de parcours au moment adéquat, soigner, etc.

Certains étudiants semblent avoir de la chance : quand on leur remet le rat dont ils auront à s'occuper, ils apprennent qu'il est issu d'une lignée particulièrement intelligente : «*c'est un rat qui devrait apprendre très vite et très bien.*» D'autres étudiants sont beaucoup moins heureux : le rat qu'ils reçoivent provient d'une lignée caractérisée par la bêtise. Les étudiants se mettent au travail et essayent de faire apprendre le labyrinthe à leur rat. Chaque jour de la session d'exercices, la performance de tous les rats est testée.

D'après les informations que nous venons de donner, les résultats du test de performance devraient être triviaux : les rats intelligents réussissent mieux que leurs congénères

6. Une illustration de cette expérience apparaît dans le film "I comme Icare" d'Henri Verneuil.

moins malins. C'est de fait ce qu'il s'est produit mais ce résultat est loin d'être aussi anodin qu'il y paraît. En effet, les rats ne proviennent pas de lignées différentes; ils avaient été distribués au hasard aux étudiants et c'est arbitrairement qu'ils avaient été qualifiés d'intelligents ou de bêtes.

Rosenthal a conduit d'autres expériences du même type qui visent toutes à mettre en évidence l'effet Pygmalion qui se traduit par une influence des attentes du sujet envers son partenaire qui ont une influence déterminante sur les résultats expérimentaux effectifs.

Ainsi, pour l'expérience portant sur les rats, l'effet Pygmalion se traduit par le fait que les étudiants qui croyaient devoir faire apprendre un labyrinthe à un rat très intelligent, le cajolaient beaucoup plus, lui parlaient davantage, etc.

À l'issue de cette expérience, on peut remarquer que c'est l'état d'esprit dans lequel les étudiants ont été placés au départ qui va jouer un rôle déterminant. Il semble que nous puissions, à cet instant de la démonstration, resituer la notion d'état d'esprit dans une problématique plus générale qui est celle du contexte. Ce contexte dans lequel chacun des étudiants s'est placé (rat intelligent ou non intelligent par exemple) a eu pour influence de modifier le résultat de l'expérience.

En conclusion, on pourra remarquer que la notion de contexte subsume bien des notions abordées en psychologie sociale. Par ailleurs, on peut remarquer que les psychologues accusent un certain retard sur les linguistes car c'est très tôt que ces derniers se sont heurtés à la notion de contexte comme cela va apparaître à présent.

1.3.2 Le contexte en pragmatique

Nombreux sont les travaux se fondant ou portant sur la notion contexte en pragmatique. Nous en donnerons la définition suivante tirée de [Moeschler et Reboul, 1994]:

Le terme de *contexte* a été utilisé avec différentes significations dans la littérature linguistique. Il est fréquemment employé pour désigner l'environnement linguistique d'un terme ou d'un énoncé, c'est-à-dire le discours où ce terme ou cet énoncé apparaît⁷. Il peut aussi désigner un ensemble d'informations dont dispose l'interlocuteur et qui lui servent pour interpréter le discours ou le fragment de discours considéré.

Sperber et Wilson [1989] l'utilisent dans la seconde acception de cette définition en insistant sur le fait que le contexte n'est pas donné une fois pour toutes, mais se construit par l'interlocuteur énoncé après énoncé sur la base d'informations provenant de sources diverses et du *principe de pertinence*.

Définir le sens d'une expression, d'une phrase ou d'un discours plus élaboré en dehors de tout contexte est quelque chose de relativement difficile. Les errements de vocabulaire concernant les concepts de sens et de signification traduisent bien les difficultés qu'ont eues linguistes et philosophes pour cerner l'ensemble des phénomènes qui régissent l'usage de la langue, en situation ou hors situation [Pierrel et Romary, 1997].

7. Dans ce cas, on parlera de co-texte.

Par ailleurs, la pragmatique fait appel au contexte comme le démontrent Reboul et Moeschler [1994]. Le contexte intervient classiquement au travers des déictiques de lieu et de personne comme “*maintenant*”, “*ici*”, “*je*” par exemple qui ne peuvent s’interpréter sans leur contexte d’énonciation. Il semble que l’on puisse multiplier à l’infini les exemples qui démontrent l’influence du contexte dans l’interprétation phrastique. Ainsi, prenons l’exemple des phrases (3) et (4) :

Le Président de la République française qui a été élu le 10 mai 1981 a été réélu en 1988. (3)

Le chat noir appartient à Philippe. (4)

Il semble intéressant de comparer la description définie qui apparaît en (3), “*le Président de la République française qui a été élu le 10 mai 1981*”, et celle qui apparaît en (4), “*le chat noir*”. La première permet de déterminer un référent, quelle que soit la situation dans laquelle elle est énoncée. Pour la seconde, la situation est bien différente : si, en effet, il y a un et un seul individu qui satisfait aux conditions spécifiées par la description définie “*le Président de la République française qui a été élu le 10 mai 1981*”, combien y a-t-il d’animaux qui satisfont les conditions spécifiées par la description définie “*le chat noir*”? Probablement une multitude. Ici, il est évident que la situation d’énonciation a un rôle à jouer : si (4) est prononcée devant une cage dans laquelle il y a trois chats, un chat noir, un chat gris et un chat roux, on aura naturellement tendance à penser que c’est le chat noir enfermé dans la cage que veut désigner le locuteur, plutôt qu’un autre animal qui ne serait pas en vue. La description définie “*le Président de la République française qui a été élu le 10 mai 1981*” sera dite *complète* puisqu’elle suffit à identifier un individu unique par opposition à la description “*le chat noir*” qui sera dite *incomplète* car elle ne saurait suffire par elle-même à identifier un unique animal. Il ressort du second exemple que l’analyse linguistique seule ne suffit pas toujours à déterminer le ou les bons référents. Il semble au contraire qu’il faille prendre en compte d’autres facteurs, généralement extra-linguistiques, pour lesquels une analyse pragmatique se révèle nécessaire.

Sperber et Wilson s’attachent à décrire plus finement ce phénomène de désambiguïsation au travers du système périphérique linguistique⁸ dont ils pensent qu’il construit effectivement les différentes interprétations de chaque constituant de la phrase et que ces interprétations sont soumises, au fur et à mesure de l’analyse linguistique, au système central qui choisit une interprétation unique (dans la mesure où il dispose des informations nécessaires). Ce système permet de réduire le coût de traitement de l’énoncé puisqu’il inhibe une partie du processus interprétatif sémantique. Cette diminution du coût du traitement nous amènera à parler au paragraphe suivant du principe d’économie cognitive.

8. Selon Fodor [1986], le traitement de l’information se fait en plusieurs étapes, dont certaines sont spécialisées alors que d’autres ne le sont pas. Les données perceptives, qui peuvent être de natures diverses (perceptions visuelles, auditives, tactiles, olfactives, linguistiques, etc.), sont d’abord traitées par des *transducteurs* dont la tâche est de les traduire pour les rendre accessibles aux mécanismes ultérieurs. Les transducteurs sont des mécanismes spécialisés et fournissent des données qui sont ensuite traitées par les *systèmes périphériques* (ou *système d’interface*) qui en offrent une première analyse. Il y a enfin le *système central* qui, lui, n’est pas spécialisé et qui travaille sur les données fournies par les systèmes périphériques (d’après [Moeschler et Reboul, 1994]).

Ainsi, pour la phrase (5) tirée de [Moeschler et Reboul, 1994], deux interprétations sont possibles : (6) et (7). La phrase (6) fournit l'interprétation la plus fréquente hors contexte :

<i>La petite brise la glace.</i>	(5)
<i>La petite fille brise la glace.</i>	(6)
<i>La petite brise lui donne froid.</i>	(7)

Sperber et Wilson mentionnent que c'est très tôt dans l'analyse linguistique de l'énoncé que l'une des deux interprétations est retenue aux dépens de l'autre. Dès lors, le système central inhibe purement et simplement la production par le système périphérique de l'autre interprétation.

Nous prendrons un dernier exemple pour illustrer les effets du contexte, et ce, au niveau spatial. Herskovits [1986] citée par Vieu [1991] mentionne très clairement l'importance du contexte: «on peut toujours trouver un contexte "marqué" dans lequel une expression locative sera admise (ou rejetée) dès lors qu'un élément de sa sémantique est un tant soit peu présent (ou absent)». Ainsi, prenons l'exemple d'un chat pour lequel il est essentiel qu'il ne marche pas sur un tapis, on dira "*le chat est sur le tapis!*" dès qu'il y met une patte. Inversement, si le chat est supposé rester sur son tapis, on dira "*le chat est hors du tapis*" dès que sa patte en dépassera.

Tout comme en linguistique, la notion de contexte a un rôle très important à jouer en informatique et surtout en dialogue homme-machine.

1.3.3 Le contexte en dialogue homme-machine

Dans le dialogue homme-machine, le contexte général de la tâche a un rôle essentiel à jouer. On définit généralement la notion de tâche comme l'objet du face-à-face homme-machine, c'est-à-dire qu'une tâche consiste pour l'opérateur en l'atteinte d'un but, ce but représentant la fin d'une étape particulière de son activité. La tâche conditionne directement les performances du système. En effet, si l'on sait réaliser de bons systèmes dans des domaines restreints, ce n'est plus vrai pour des domaines plus vastes car on assiste très vite à une explosion combinatoire des interprétations possibles [Carré *et al.*, 1991].

De notre côté, le contexte aura une influence déterminante sur le modèle de calcul des référents que nous élaborerons au chapitre 3. Il faudra que notre méthode de calcul soit valable dans tous les contextes en s'adaptant finement à ces derniers. Pour illustrer ce dernier point, prenons l'exemple de la phrase (8) qu'un utilisateur énoncerait tour à tour devant l'écran 1 puis devant l'écran 2 de la figure 1.1 où A et B désignent des fenêtres⁹ et X, Y, Z désignent des icônes¹⁰.

9. Fenêtre : au sens informatique du terme, il s'agit d'un rectangle représentant une fenêtre virtuelle qui peut contenir du texte, des images, des boutons, etc.

10. Icône : symbole ou image réduite d'une fenêtre qu'elle s'attache à représenter. Son principal rôle est de limiter l'encombrement au niveau de l'écran. Un clic sur une icône rétablira la fenêtre d'origine, de

Détruis l'icône qui se trouve à droite de la fenêtre A. (8)

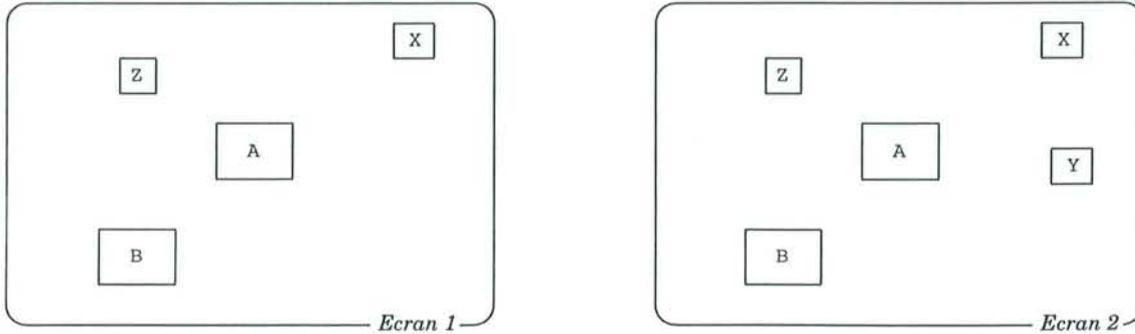


FIG. 1.1 – Prépositions spatiales et contexte.

Si c'est l'icône X qui sera détruite dans le cas de l'écran 1, en revanche, pour l'écran 2, il s'agira de l'icône Y. Ainsi, alors que la fenêtre A et l'icône X occupent la même position dans les deux écrans, c'est le contexte environnant la fenêtre A qui a changé avec l'apparition de l'icône Y à l'écran. Ce nouveau contexte conduit à changer l'interprétation de la préposition "à droite de".

Dans une perspective d'implémentation, la prise en compte du contexte pourrait nous conduire à l'introduction de plusieurs algorithmes afin de tenir compte de certains contextes types pour aboutir alors à la définition de plusieurs prépositions "à droite de" comme cela a pu apparaître dans les travaux de Hernández [1994]. Nous pensons que cette solution n'est pas souhaitable et lui préférons, dès que cela sera possible, une définition unifiée qui s'adaptera à tel ou tel contexte.

Enfin, si l'on peut considérer à première vue, dans l'exemple précédent, que le contexte environnant A recouvre tout l'écran, il n'en va pas toujours ainsi. Comme nous le démontrerons dans le chapitre 4, concernant les cadres de référence, ces derniers, de par leur fonction à considérer l'ensemble des entités nécessaires et suffisantes à l'interprétation de la phrase courante, posséderont une étendue qui n'occupera bien souvent qu'une sous-partie de l'écran.

1.3.4 Conclusion

Comme nous venons de le voir, la notion de contexte joue un rôle déterminant que ce soit en psychologie, en pragmatique ou en informatique. En ayant eu recours à la psychologie et plus particulièrement à la psychologie sociale, nous avons pu éclairer la notion de contexte sous un angle différent de celui auquel l'informaticien est habitué.

même qu'un clic sur une partie spécifique de la fenêtre verra cette dernière disparaître de l'écran pour laisser place à son icône.

Du côté de l'informatique et plus particulièrement du point de vue de la réalisation d'une maquette informatique, la notion de contexte¹¹ sera omniprésente tout au long de cette thèse. Elle nous guidera dans l'élaboration de nos algorithmes qui seront paramétrés suivant le contexte.

Si l'on se place dans une perspective linguistique, nous avons pu esquisser que la notion de cadre s'apparentera largement à ce que l'on pourra dénommer un contexte spatial. Nous insisterons sur ce point au paragraphe 4.3.1.

Enfin, il semble que le contexte puisse jouer un rôle décisif dans l'allégement de la tâche cognitive d'un sujet humain. Ce point va nous amener à parler tout naturellement de principe d'économie cognitive au paragraphe suivant.

1.4 Principe d'économie cognitive

De manière plus générale, on peut se demander si les notions de dialogue et de contexte ne sont pas gouvernées par des principes plus généraux. De notre point de vue, le principe d'économie cognitive semble faire partie de l'un d'entre eux. Voici la définition que nous en proposons :

Principe d'économie cognitive (p.é.c dans la suite): principe illustrant la diversité des mécanismes mis en œuvre par un sujet donné en vue d'alléger sa charge cognitive.

Nous proposons de définir la notion de charge cognitive comme suit :

Charge cognitive: énergie cérébrale déployée par un cerveau humain telle qu'elle pourrait être mise en évidence par un tomographe par émission de positons.

Le bien fondé de cette dernière définition apparaîtra dans les travaux de Haier que nous présenterons ci-après.

Nous intéressant tout d'abord à certains aspects du fonctionnement humain, nous tenterons de démontrer que le p.é.c semble en mesure d'expliquer certaines facettes de la cognition humaine. Dans un second temps, nous tenterons de montrer que le p.é.c sous-tend également une partie de la pragmatique. Dans un troisième temps, en vue de créer des systèmes informatiques dont le comportement se rapprocherait le plus possible du fonctionnement d'un sujet humain, il nous semble essentiel de s'inspirer du p.é.c. En effet, ce principe est de nature à fournir des heuristiques très intéressantes en vue d'une implémentation informatique. Enfin, dans la dernière partie, nous examinerons la notion de p.é.c sous un angle critique. Si cette notion semble intuitivement très satisfaisante, nous verrons que son application est parfois très délicate. En effet, encore plus générale que la notion de pertinence, la notion de p.é.c est parfois difficile à saisir dans les faits.

11. Nous parlons ici de la notion de contexte dans son acception unifiée, en effet, ainsi que cela est apparu ci-dessus, on peut difficilement mentionner d'une unique notion de contexte.

1.4.1 Le principe d'économie cognitive à la base de la cognition humaine ?

Pour tenter d'apporter quelques éléments de réponse à cette vaste question, nous emprunterons de nombreux exemples provenant principalement de la psychologie sociale.

a) P.é.c et mémoire

Nous nous pencherons dans ce paragraphe plus spécifiquement sur la mémoire à court terme que l'on oppose à la mémoire à long terme. Nous rappelons ici d'après Beaugé [1995] les principales caractéristiques qui les opposent :

caractéristiques	mémoire à court terme	mémoire à long terme
durée de rétention	1 à 3 min	quelques heures à plusieurs années
capacité	limitée empan de 7 items	supposée illimitée
fréquence de renouvellement	labile	stable consolidée
sensibilité aux interférences	extrême	faible

L'homme est plongé dans un monde qui le submerge de plus en plus d'informations, et pourtant, sa capacité à traiter ces informations reste limitée. Miller [1956] cite à ce propos le nombre magique de *sept* informations que nous pourrions traiter simultanément : lorsque nous avons à mémoriser à court terme une série de lettres, nous sommes capables de le faire pour environ sept lettres (en fait 5 à 9 lettres, suivant les cas). La performance reste identique lorsque, plutôt que des lettres, il s'agit de mémoriser des syllabes dépourvues de sens, ou des mots. Nous emmagasinons l'information jusqu'à concurrence de sept paquets. Qu'il s'agisse de lettres isolées, de syllabes ou de mots, cela fait toujours sept paquets : le nombre de lettres mémorisées augmente mais le nombre de catégories reste identique. La principale conséquence de cette limitation est qu'elle nous oblige, nous autres humains, à structurer sans cesse l'espace qui nous entoure compte tenu de la quantité d'objets placés dans notre environnement et de nos faibles capacités de mémorisation.

Nombre de "recettes de cuisine" ont d'ailleurs vu le jour afin de nous permettre de retenir facilement des listes d'objets ou des listes de nombres quelconques sans trop d'efforts. Ainsi, ces efforts se voient souvent atténués en procédant par exemple à l'aide d'associations d'idées. Ces associations opèrent soit entre les objets à mémoriser, soit entre ces derniers et des objets bien connus du sujet humain ou encore insolites.

b) P.é.c et catégorisation

L'opération de *catégorisation*¹² nous semble être l'outil fondamental propre à faciliter l'allègement de la charge cognitive. Ainsi, catégoriser permet de savoir ou de dire beaucoup de choses à partir de peu d'éléments, et d'apprendre ou de retenir, peu de choses à partir de beaucoup d'éléments. Rosch [1978] a également insisté sur ce point en indiquant que la catégorisation nous permet de réduire la complexité de notre environnement.

c) P.é.c et scripts

Une autre illustration de la notion de charge cognitive et donc du p.é.c peut être mise en avant à l'aide de la notion de *script* que Leyens définit comme suit :

Script : «séquence cohérente d'événements attendus par l'individu qui l'impliquent soit comme participant, soit comme observateur» [Leyens, 1992].

Comme le mentionne Leyens, les scripts sont éminemment utiles dans la vie quotidienne. Ce sont des catégories d'événements qui nous permettent de répondre à de larges tranches d'informations sans trop d'efforts, sans y prendre garde lorsque le script est bien mis au point, ce qui nous rend disponibles pour d'autres tâches. Notre fonctionnement cognitif devient alors le suivant : lorsqu'une activité se répète fréquemment, elle devient un script à la structure duquel nous restons sensibles alors que nous ne faisons plus attention à son contenu sémantique. Face à un tel script, nous agissons sans trop réfléchir.

Si par contre, la structure n'est pas typique de celle attendue, le contenu sémantique reprendra de l'importance ; il s'agira, en effet, de voir précisément ce dont il est question.

De même, si la situation demande un effort particulier de notre part, le script ne sera pas traité automatiquement car son contenu sémantique importera aussi¹³.

d) P.é.c et gestaltisme

Selon nous, la Gestalt Wertheimer [1912], Koffka [1935] et Kohler [1947] se fonde également sur le p.é.c. Pour mémoire, la psychologie de la Gestalt, apparue en 1912, est basée sur l'idée que le tout est différent de la somme des parties. Cette théorie concurrença la théorie du structuralisme alors prédominante. Ainsi, l'un des concepts d'organisation du gestaltisme relève du principe de simplicité selon lequel, lorsque les stimuli sont ambigus, la perception choisira la bonne forme en cherchant celle qui est simple, régulière et symétrique, qui apparaît comme dominante.

e) P.é.c et imagerie mentale

Il semble que le p.é.c ait été mis en lumière au niveau du cerveau lui même comme le démontrent les études conduites au centre d'imagerie cérébrale de l'Université de Californie par le Pr Richard Haier et que relate E. Pigani [1995]. Haier a demandé à des volontaires de jouer avec un puzzle vidéo alors qu'ils se trouvaient sous l'observation d'un

12. Pour une bonne synthèse du domaine, voir l'ouvrage de Dubois [1991b].

13. On raconte que, dans l'obligation de subvenir à ses besoins, Einstein avait choisi le travail le plus machinal possible afin de pouvoir, tout en travaillant, penser à des problèmes théoriques de physique.

T.E.P¹⁴. Haier a découvert que les patients déployaient une énergie cérébrale considérable dans la phase d'apprentissage du jeu, laquelle diminuait après plusieurs jours de pratique. D'après les résultats de cette expérience, moins un étudiant dépensait d'énergie cérébrale, plus son quotient intellectuel était élevé. Haier en a conclu que l'intelligence serait probablement une question d'efficacité du cerveau : un cerveau intelligent cherche à travailler moins en mettant en action moins de neurones, ou moins de circuits, ou moins des deux. À l'inverse, un cerveau moins intelligent fait appel à un grand nombre de circuits neuronaux non concernés ou inefficaces...

f) P.é.c et la dualité des modèles qualitatifs-quantitatifs

Pour étayer notre démonstration, nous prendrons un dernier exemple tiré de l'aptitude humaine à raisonner plutôt de manière *qualitative* que *quantitative*. Alors que le raisonnement quantitatif impose très vite une lourde charge cognitive [Hernández, 1994], le raisonnement qualitatif permet au contraire de s'affranchir des détails inutiles en ne mentionnant, dans un contexte donné, que les aspects nécessaires qui autoriseront la discrimination voulue parmi l'ensemble des alternatives possibles. À titre d'exemple on peut prendre le cas d'un système de multifenêtrage pour lequel un utilisateur préférerait sans doute dire "*mets la fenêtre verte à droite*" s'il veut placer cette dernière à droite de son écran, plutôt que "*déplace la fenêtre verte de 10 cm vers la droite*".

Comme cela sera mis en valeur au prochain paragraphe, il semble que la notion de p.é.c puisse subsumer certains travaux effectués en pragmatique.

1.4.2 Le principe d'économie cognitive en pragmatique

Selon nous, la notion de p.é.c trouve sa meilleure illustration en pragmatique au sein des travaux de Grice [1979] concernant les maximes de conversation et ceux de Sperber et Wilson [1989] concernant la notion de pertinence.

Selon Grice, dans un échange conversationnel, les contributions des locuteurs sont gouvernées par un principe général qu'il nomme le principe de coopération.

Il explicite ce principe suivant quatre catégories générales qui sont appelées les maximes de conversation et que nous rappelons ici :

- **Maximes de quantité**
 1. Que votre contribution contienne autant d'information qu'il est requis.
 2. Que votre contribution ne contienne pas plus d'information qu'il n'est requis.
- **Maximes de qualité (de véridicité)**

Que votre contribution soit véridique :

 1. N'affirmez pas ce que vous croyez être faux.
 2. N'affirmez pas ce pour quoi vous manquez de preuves.

14. T.E.P : tomographe par émission de positons.

- **Maxime de relation (de pertinence)**

Parlez à-propos (soyez pertinent).

- **Maximes de manière**

Soyez clair :

1. Évitez de vous exprimer avec obscurité.
2. Évitez d'être ambigu.
3. Soyez bref (évitez toute prolixité inutile).
4. Soyez ordonné.

Si la maxime de quantité se situe directement dans la veine du p.é.c, les trois autres maximes contribuent également, à des degrés divers, dans ce sens : ainsi, il paraît évident qu'une expression obscure ou ambiguë nécessitera plus d'efforts pour être comprise qu'une qui ne le serait pas.

Les travaux ultérieurs de Sperber et Wilson [1989] sur la pertinence semblent s'intégrer également dans une perspective d'allègement cognitif suivant le concept *d'effort de traitement* tel qu'il apparaît dans cette définition de la pertinence :

Pertinence :

- (a) Toutes choses étant égales par ailleurs, plus un énoncé produit d'effets contextuels, plus cet énoncé est pertinent.
- (b) Toutes choses étant égales par ailleurs, moins un énoncé demande *d'efforts de traitement*, plus cet énoncé est pertinent.

La notion de p.é.c semble donc en mesure d'unifier la théorie de Grice ainsi que celle de Sperber et Wilson. De plus, si ces deux théories sont très utilisées en dialogue homme-machine, il semblerait que le p.é.c puisse également jouer un rôle très important dans ce dernier domaine.

1.4.3 Application au niveau du dialogue homme-machine

Sans faire une taxinomie complète de l'ensemble des tâches qui peuvent alourdir la charge cognitive d'un sujet humain placé en situation d'interaction avec sa machine, nous nous attarderons sur celle qui semble la faire augmenter de manière très significative selon que l'utilisateur se pose plus ou moins la question de savoir : "Comment vais-je traduire ma requête afin que l'ordinateur la comprenne?"

Nous savons que dans un dialogue homme-homme, nous transmettons l'information dans la majorité des cas en fonction de ce qu'on imagine que l'interlocuteur va comprendre. Si nous appliquons ce principe au dialogue homme-machine, en sachant que nous sommes encore loin de pouvoir disposer d'interfaces homme-machine "naturelles", cela obligera l'utilisateur à se forger au fil du dialogue une image des facultés et inaptitudes potentielles qui lui sont offertes par son ordinateur, ce qui alourdira sa charge cognitive.

Corrélativement, plus les facultés de compréhension de l'ordinateur seront grandes, moins l'utilisateur aura d'efforts à fournir pour formuler ses requêtes. L'idéal vers lequel nous souhaitons tendre consiste en un "dialogue naturel" dans lequel l'utilisateur ne se focalisera plus sur la question de savoir "comment" traduire sa requête mais beaucoup

plus sur la tâche qu'il souhaite réaliser.

C. Faure et M. Arnold [1993] illustrent plus précisément ce dernier point dans la mesure où elles démontrent que, dans une interaction, la manière dont une information est transmise résulte de la combinaison des trois facteurs principaux d'économie suivants : la rapidité, la précision, la simplicité : la rapidité concerne le temps d'émission (de réception) du message ; la précision vise à produire des messages non-ambigus et la simplicité d'un message relève de la complexité des mécanismes nécessaires à la construction (à l'analyse) de son expression. Les effets de ces facteurs peuvent s'opposer dans une recherche d'économie qui doit donc optimiser leur combinaison. Par exemple, la commande verbale la plus courte comme "déplace le" demande une recherche du référent, soit dans l'historique des commandes, soit dans des canaux parallèles de réception d'information, ce qui joue sur la complexité du processus d'interprétation de l'énoncé.

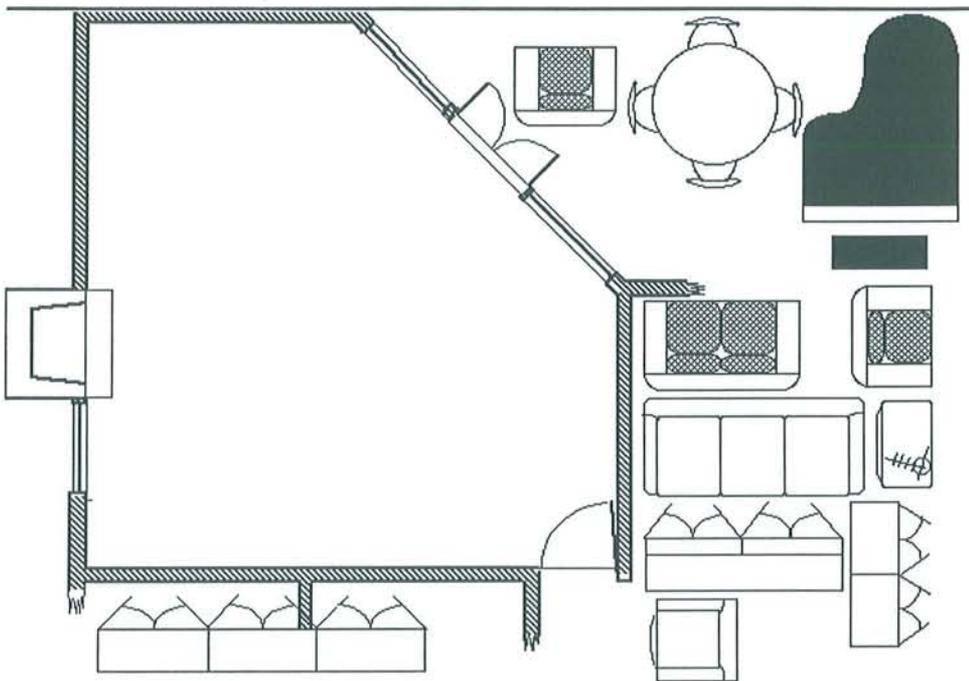


FIG. 1.2 – Visualisation d'un salon pour une application d'aménagement.

Faure et Arnold mentionnent également que l'évolution des interfaces homme-machine passera nécessairement par une utilisation accrue de commandes "incomplètes", c'est-à-dire des énoncés linguistiques en situation qui ne contiennent pas explicitement toute l'information nécessaire au lancement d'une action mais qui peuvent être complétés par des inférences contextuelles. Les commandes "incomplètes" permettent bien entendu un surcroît de rapidité et de simplicité, le facteur de précision étant réduit au maximum par le recours au contexte. Dans une perspective plus large, on peut remarquer que nous touchons ici à la base du fonctionnement linguistique.

Au niveau de la conception informatique, il semblerait que cet allègement cognitif passe par la réalisation d'interfaces multimodales comme le démontrent les travaux qu'a réalisés Mignot dans notre équipe [Mignot *et al.*, 1993] et [Mignot, 1995].

L'expérience de type Magicien d'Oz qu'il a conduite en collaboration avec le CERMA¹⁵ [Dauchy *et al.*, 1993] avait pour but d'étudier le comportement de sujets dans une situation de dialogue multimodal (associant voix et geste). L'un des espaces de la tâche tel qu'il pouvait être vu dans cette expérience est représenté à la figure 1.2. Il comprend d'une part une portion d'appartement (un salon...), avec une fenêtre, une porte et une cheminée, et d'autre part un ensemble de meubles disposés à l'extérieur de la pièce qu'il s'agit de positionner (en partie) dans celle-ci. On peut imaginer que ces meubles sont en quelque sorte en vrac dans un camion de déménagement et que l'utilisateur, joue le rôle d'un déménageur, les met en place un à un. Mignot propose une classification¹⁶ en six types d'interactions possibles (voir figure 1.3) :

- V : l'énoncé oral seul suffit à réaliser la requête.
- V ou G : l'énoncé oral seul aurait pu suffire à réaliser la requête, il en va de même du geste.
- Gv : l'énoncé oral est redondant par rapport au geste, le geste seul aurait suffi à réaliser la requête.
- Vg : Idem requête précédente en échangeant le rôle du geste et de l'énoncé oral.
- V et G : l'énoncé oral et le geste se complètent parfaitement, ils servent tous les deux à la réalisation de la requête.
- G : geste seul.

Parmi les six sujets ayant servi à cette expérimentation, deux n'utilisaient que des messages purement vocaux, deux utilisaient indifféremment le geste et la parole, les deux derniers étaient purement gestuels. Mignot a conclu de ces expériences que les interfaces multimodales semblent, de par leur grande souplesse, pouvoir s'adapter aux variations intra et inter-individuelles. Qui plus est, la multimodalité semble pouvoir améliorer la concision du dialogue et donc simplifier la tâche de l'utilisateur en allégeant sa charge cognitive. Ainsi, il semble naturel de penser qu'une majorité d'utilisateurs préférera utiliser la requête (10) plutôt que la requête (9) pour détruire une icône qui se trouve affichée à l'écran¹⁷ :

Détruis l'icône qui se trouve en haut à droite de la fenêtre verte. (9)

Détruis cette icône. (+ clic souris) (10)

Dans la suite de cette thèse, le p.é.c s'appliquera principalement dans les cas suivants :

- Nous avons pu voir dans ce chapitre que les possibilités de mémorisation spontanées d'un être humain se "cristallisaient" aux alentours de la valeur 7 ce qui a pour effet de limiter considérablement ses performances. Pour le cas de l'aménagement du salon que nous retrouverons au chapitre 4, nous verrons que cela oblige l'homme à structurer la scène qu'il perçoit. Nous verrons dans ce même chapitre un modèle

15. CERMA : Centre d'Études et de Recherches de Médecine Aérospatiale.

16. Il mentionne clairement que ce n'est pas la seule possible.

17. Bien entendu si l'utilisateur n'est pas novice, il préférera sans doute cliquer directement sur le bouton qui permet de détruire l'icône.

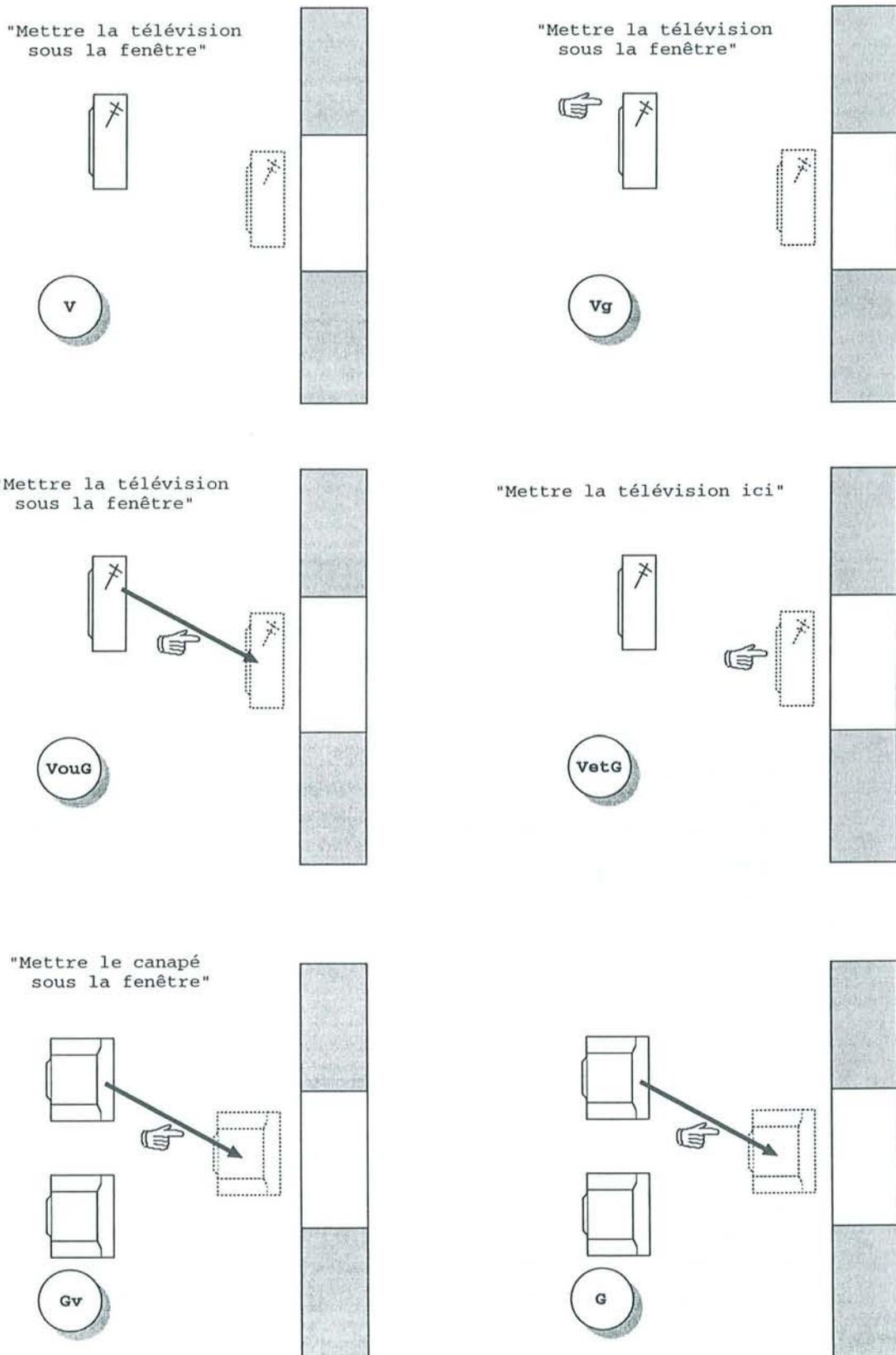


FIG. 1.3 – Une classification possible des commandes multimodales vocales et gestuelles.

informatique qui simule le raisonnement spatial humain en proposant une structuration de la scène qui se trouve affichée à l'écran.

- Dans la perspective de réaliser un module de calcul des référents spatiaux, nous avons pu démontrer dans [Schang, 1991] que les performances du système réalisé dépendaient fortement de l'ensemble des objets qu'il fallait prendre en compte à un instant donné du dialogue. Pour éviter une éventuelle explosion combinatoire, il semble donc souhaitable de restreindre autant que possible cet ensemble aux seuls objets pertinents. Nous tenterons de démontrer que la notion de cadre semble être un bon outil pour permettre de restreindre cette combinatoire du fait que tout ce qui se trouvera, à un instant donné du dialogue, à l'extérieur du cadre, ne sera pas pris en compte.
- Enfin, lors de l'élaboration du modèle de calcul des référents du chapitre 4, nous chercherons à réaliser un modèle qui soit le plus simple possible. Nous préférons développer un modèle plutôt qualitatif que quantitatif car il nous semble primordial d'éviter autant que possible le recours à des formules mathématiques complexes. Nous justifierons ce dernier point par le fait que l'humain raisonne bien au jour le jour en s'affranchissant de calculs complexes (tout du moins, si nous nous référons à la connaissance que nous avons du fonctionnement de ce raisonnement humain tel que nous en sommes conscients "en surface"). Nous nous plaçons ici dans une perspective de raisonnement de haut niveau, par opposition aux domaines de l'intelligence artificielle qui requièrent l'élaboration de ce que Fodor appellerait les transducteurs. Ces derniers relèvent de l'intelligence artificielle dite de "bas niveau" tels, par exemple, la reconnaissance des formes. Au sein du bas niveau, on ne sait pas proposer de modèles "simples" qui possèdent par exemple une inspiration purement qualitative (voir à titre d'illustration, les travaux de Haton *et al.* [1991] concernant le problème de la reconnaissance automatique de la parole).

Au terme de ce paragraphe, on peut sans doute croire que la notion de p.é.c est suffisamment générale pour s'appliquer partout, ou presque, en informatique. Il semble cependant qu'il faille nuancer la portée de ce principe et le paragraphe suivant va apporter quelques éléments de réponse à ce propos.

1.4.4 Réflexion sur l'application du principe d'économie cognitive

Au vu de ce qui précède, le risque est peut-être grand de vouloir appliquer le principe d'économie cognitive partout. Il semble toutefois que son application soit parfois délicate. Ainsi, prenons le problème spécifique de la catégorisation au sein des langues dont nous rappelons ici brièvement l'évolution. La question à laquelle ont cherché à répondre les philosophes depuis l'antiquité était la suivante: «*sur quels critères peut-on décider de l'appartenance ou de la non appartenance d'un objet à une catégorie?* ». Leur réponse a été la suivante: «*les membres d'une même catégorie partagent les mêmes caractéristiques et le critère d'appartenance à la catégorie en question passe par la possession de ces*

caractéristiques ». Dès lors, la catégorisation repose sur un modèle à base de conditions nécessaires et suffisantes :

- les catégories sont des entités aux frontières clairement délimitées.
- l'appartenance d'une entité particulière à une catégorie répond au système du vrai ou faux : un X quelconque est un "chien" ou n'est pas un "chien", selon qu'il satisfait ou non aux conditions critères de la catégorie "chien".
- les membres d'une même catégorie ont un statut catégoriel égal, puisque chaque membre possède les propriétés requises par la définition de la catégorie. Eu égard à la catégorie à laquelle ils appartiennent, chaque membre est un aussi "bon" membre que les autres.

Les travaux de Rosch [1978] ont démontré par la suite que les langues ne peuvent être découpées en *catégories classiques* mais qu'il faut au contraire recourir aux *catégories naturelles*¹⁸. Ces dernières n'ont pas de frontières bien définies, il n'existe pas nécessairement une propriété partagée par tous les membres de la catégorie, les frontières sont variables et dépendent des modifications du système entier, enfin, tous les membres de la catégorie n'ont pas le même statut, l'usage (ou les usages) le(s) les plus représentatif(s) est(sont) appelé(s) *prototype(s)*.

Au premier abord, le p.é.c semble bien mieux s'inscrire dans le modèle fonctionnant en termes de conditions nécessaires et suffisantes que celui issu de la théorie standard. En effet, la théorie standard semble bien plus complexe et floue.

Comment expliquer alors que la cognition humaine, dont nous avons postulé quelle était sous-tendue par le p.é.c, s'écarte aussi sensiblement de principes simples pour leur préférer finalement ceux de la théorie standard qui paraissent bien plus complexes ? Il semble que la réponse ne soit pas aussi simple. Tout d'abord, il est très difficile, voire impossible pour l'instant, de cerner tous les phénomènes cognitifs qui interviennent au sein de la catégorisation. En particulier, la mémoire semble avoir un rôle décisif à jouer, or la compréhension de son fonctionnement n'en est pour l'instant qu'à ses premiers balbutiements [Beaugé, 1995].

D'autre part, on peut se demander si le modèle fondé sur les conditions nécessaires et suffisantes est aussi économe que prévu d'un point de vue cognitif ? En particulier, si l'on s'interroge sur l'appartenance du mot "poussin" à la catégorie "oiseau", il serait pour le moins coûteux d'un point de vue cognitif de balayer toutes les conditions nécessaires et suffisantes pour décider à la fin si oui ou non "poussin" fait partie de la catégorie "oiseau". On peut dès lors concevoir qu'un fonctionnement cognitif à base de prototypes puisse être plus économe.

1.4.5 Conclusion

Ces différents points nous confortent dans le fait que le p.é.c reflète bien l'un des aspects fondamentaux du fonctionnement du système cognitif humain dans le sens où ce dernier cherche à être le plus efficace possible, allant dès que possible à l'essentiel. Par ailleurs, si la notion de principe d'économie cognitive ne parvient pas à expliquer tous

18. Cette théorie porte le nom de "théorie standard" par opposition à la "théorie étendue". Voir [Kleiber, 1990a] pour un bon descriptif de ces travaux ou [Dubois, 1991a] et [Kleiber, 1991] pour un résumé détaillé.

les phénomènes de la cognition humaine, nous avons pu voir que ce principe parvenait à donner un point de vue commun sur les maximes de Grice et la notion de pertinence de Sperber et Wilson.

Dans une perspective d'implémentation, le p.é.c sera sous-jacent à tous les modèles que nous allons proposer. En particulier, le p.é.c fait partie intégrante de la notion de cadre.

1.5 Illustration du tryptique dialogue, contexte et principe d'économie cognitive

Les travaux de Garrod [1987] et [1994] nous confortent dans nos dires quant à l'importance des trois notions abordées dans ce chapitre.

Garrod étudie la manière dont les participants à une conversation coordonnent leur utilisation et leur interprétation du langage dans un contexte restreint. En particulier, son étude repose sur l'analyse de descriptions spatiales qui ont été obtenues en laboratoire sur la base d'un jeu de labyrinthe qui apparaît à l'écran d'un ordinateur. Le principe de l'expérience est le suivant : on prend deux joueurs A et B que l'on place dans deux salles différentes, chacun étant placé devant un écran d'ordinateur sur lequel se trouve affiché un labyrinthe (voir la figure 1.4.(a) pour l'écran du joueur A). Les deux joueurs jouent contre l'ordinateur et gagnent lorsqu'ils réussissent à atteindre une position prédéterminée du labyrinthe. Par ailleurs, certains tunnels sont fermés (voir figure 1.4), le seul moyen de les ouvrir est d'amener l'autre joueur sur une zone de "switch" qui permet alors de changer les tunnels fermés en tunnels ouverts et réciproquement. Pour augmenter la difficulté, certaines spécificités ont été ajoutées. Tout d'abord, chacun d'eux ne voit que son propre personnage et son propre but. Chaque joueur peut indiquer à l'autre sa position comme le démontre l'extrait de corpus suivant qui a été recueilli pour les figures 1.4.(a) et 1.4.(b) :

A: *Je me trouve sur la dernière ligne en bas.*

B: *Quelle colonne ?*

A: *La troisième.*

B: *Moi, c'est la quatrième ligne en partant d'en haut, deuxième colonne.*

Enfin, pour augmenter la difficulté, un monstre cherche à attraper le personnage qui lui est le plus proche.

Garrod a effectué deux types d'analyses sur ces corpus. Tout d'abord, il a réalisé une analyse sémantique des différents types de descriptions qui indique comment les couples de locuteurs développent différents schémas linguistiques associés à différents modèles mentaux de la configuration du labyrinthe. La seconde partie de son analyse porte sur la manière dont les communicants parviennent à coordonner leurs descriptions.

De nombreux résultats ressortent de cette étude. Nous retiendrons tout d'abord ceux qui valident l'intérêt du tryptique dialogue, contexte et principe d'économie cognitive.

Côté dialogue, Garrod a bien mis en évidence le fait que les deux locuteurs coopèrent

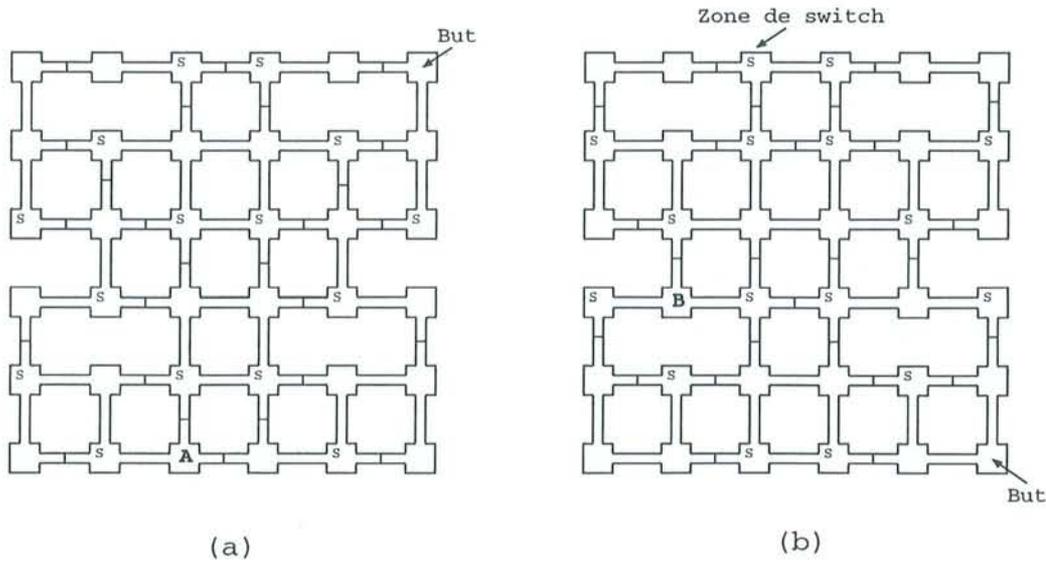


FIG. 1.4 – Exemple de deux labyrinthes pour deux joueurs A et B.

vers la réalisation d'un même but. La notion de but est essentielle en dialogue homme-machine car elle contribue à cerner l'intention du locuteur. Elle guide alors l'ordinateur vers l'ensemble des actions à réaliser en vue d'élaborer et de réaliser la requête de l'utilisateur. En effet, à la différence du domaine relatif à la compréhension de textes, le domaine spécifique du dialogue requiert de la part de l'interlocuteur (ici l'ordinateur) de réaliser une ou plusieurs actions qui passent nécessairement par la prise en compte du but que s'est fixé l'utilisateur.

Par ailleurs, Garrod a mentionné que cette collaboration entre les deux joueurs s'appuyait sur un ensemble de connaissances partagées. La notion de connaissances partagées va nous permettre de nous intéresser à présent à la notion de contexte.

Cette dernière trouve de notre point de vue, sa principale illustration dans les différents descriptifs qui sont utilisés par les joueurs. Garrod dénombre les quatre classes suivantes :

- les *path descriptions*¹⁹ qui se fondent sur un chemin virtuel qu'énonce le locuteur, exemple : "regarde le coin inférieur droit. Va deux cases à gauche et deux cases au dessus, c'est là que je suis".
- les *figural descriptions*²⁰ qui se fondent sur l'identification de configurations de nœuds particuliers, exemple : "regarde le rectangle qui se trouve en bas à gauche. Je me trouve dans sa partie inférieure gauche".
- les *co-ordinate descriptions*²¹ qui s'inspirent de la localisation qui est utilisée aux échecs, exemple : "je me trouve en C4". Ce type de localisation suppose que les deux joueurs se soient mis d'accord auparavant dans le dialogue.
- les *line descriptions*²² qui se fondent sur le découpage en lignes et colonnes du

19. Les descriptions sous forme de chemin.

20. Les descriptions figuratives.

21. Les descriptions à base de coordonnées.

22. Les descriptions se fondant sur les lignes et colonnes.

labyrinthe, exemple : “*je me trouve sur la dernière ligne, quatrième colonne*”.

L’analyse du corpus a permis à Garrod de souligner que les locuteurs se placent très rapidement dans un schéma descriptif de référence donné. Ce schéma qui peut évoluer au fil du dialogue joue le rôle de contexte commun aux deux locuteurs.

Dans une optique plus large, Garrod mentionne que ses travaux paraissent suggérer que le traitement du langage au cours d’un dialogue est régi par des principes locaux d’interaction qui ont reçu peu d’attention de la part des psychologues et des linguistes jusqu’à aujourd’hui.

De leur côté, Clark et Murphy²³ soulignent que la notion de dialogue est intimement gouvernée par la co-construction d’un ensemble de connaissances partagées. Ce dernier qui s’instaure entre les deux partenaires du dialogue permet d’éviter ou de lever les ambiguïtés référentielles.

Pour ce qui nous concerne, nous chercherons à cerner la notion de terrain commun à l’aide du concept de cadre qui jouera le rôle de contexte spatial commun à l’homme et à la machine. Ainsi, dès lors que l’ordinateur ne comprend pas une requête de l’utilisateur (lorsqu’il y a ambiguïté par exemple), il faut qu’il change de perspective (de *cadre* pour nous), jusqu’à aboutir à une unique interprétation possible.

Enfin, la notion de principe d’économie cognitive trouve au moins deux illustrations au sein des travaux de Garrod. La première illustration procède du fait que lorsqu’un utilisateur s’est placé par exemple dans une logique de description de type *path description*, il ne va pas brutalement passer à une description de type *coordinate description* par exemple. La seconde illustration découle de la notion de connaissance partagée. Garrod énonce le principe suivant «*aussi longtemps que vous pouvez interpréter la phrase de votre interlocuteur avec votre propre contexte, conservez ce dernier.*»

En conclusion, cette étude nous semble essentielle car elle démontre l’articulation forte entre les trois notions que sont le dialogue, le contexte et la notion de principe d’économie cognitive.

1.6 Synthèse

Ce chapitre définit les fondements de cette thèse au travers des trois piliers que sont la notion de dialogue, de contexte et enfin celle de principe d’économie cognitive. Outre le caractère fédérateur de chacune de ces notions au sein de son propre domaine, nous avons pu voir l’importance de chacune d’elles au sein des sciences cognitives. Par ailleurs, nous avons insisté sur la forte imbrication de ces notions entre elles. On a ainsi pu voir qu’il était difficile de parler de l’une sans glisser inmanquablement vers les autres.

23. Citation de J. Vivier [1997] : [Clark et Murphy, 1982] H. Clark and G.H. Murphy 1982, La visée vers l’auditoire dans la signification et la référence, *Bulletin de psychologie*, XXXV, 356, p. 767-777.

Ce chapitre nous a également permis de mettre en évidence les apports du tryptique dialogue, contexte et principe d'économie cognitive. En particulier, du côté du dialogue, deux résultats semblent importants. Premièrement, le dialogue homme-machine possède des spécificités que n'a pas le dialogue homme-homme. Deuxièmement, l'intention du locuteur et la vision qu'il a des performances de la machine sont centrales à l'interaction qu'il aura avec cette dernière.

La notion de contexte a été pour nous l'occasion d'étudier certains travaux effectués en psychologie sociale. Tout comme la linguistique, elle semble avoir été confrontée aux problèmes contextuels. L'expérience de Milgram ainsi que le fameux effet Pygmalion abondent dans ce sens. Du côté de l'informatique et plus particulièrement de la modélisation spatiale, nous avons pu souligner qu'il conviendra de gérer au mieux la notion de contexte.

La dernière partie de ce chapitre a été pour nous l'occasion d'aborder la notion de principe d'économie cognitive. Ainsi, de la neurologie à la Gestalt, en passant par la pragmatique au travers des travaux de Grice, de Sperber et Wilson, le principe d'économie cognitive semble être présent partout. Malgré les aspects très généraux que recouvre cette notion, ce qui la rend difficile à appliquer dans la pratique, son caractère unificateur nous incitera cependant à proposer au chapitre 4 un modèle de calcul des référents spatiaux s'en inspirant.

Par ailleurs, nous avons relaté les travaux de Garrod qui nous semblent essentiels car peu de travaux concernant les problèmes de dialogue ont trait à l'espace et, inversement, rares sont les travaux traitant de l'espace qui portent également sur des aspects dialogiques.

Enfin, la notion de cadre que nous développerons par la suite se fondera sur le tryptique qui a été étudié dans ce chapitre.

Chapitre 2

Modélisation des usages spatiaux de la langue

Sommaire

2.1	Introduction	46
2.2	Limites d'une approche géométrique	47
2.2.1	L'exemple de " <i>près de/loin de</i> "	47
2.2.2	L'exemple de la préposition " <i>dans</i> "	48
2.2.3	Conclusion	49
2.3	La prise en compte des aspects fonctionnels : une nécessité	49
2.3.1	La compositionnalité	49
2.3.2	La transitivité	50
2.3.3	Conclusion	50
2.4	Le modèle de Vandeloise	50
2.4.1	Prolégomènes : les notions de ressemblance de famille et d'impulsion	51
2.4.2	La prise en compte de la pragmatique chez Vandeloise	52
2.4.3	Les aspects fonctionnels	54
2.4.4	Un exemple : les prépositions " <i>sur/sous</i> "	56
2.4.5	Conclusion	58
2.5	Le modèle de Asher et Sablayrolles	59
2.5.1	Introduction	59
2.5.2	Taxinomie des verbes de mouvement	59
2.5.3	Sémantique des verbes de changement de lieu	61
2.5.4	Conclusion	64
2.6	Le modèle de Hernández	64
2.6.1	Les bases du modèle de Hernández	65

2.6.2	Les relations topologiques	65
2.6.3	Les relations d'orientation	66
2.6.4	Prise en compte des objets possédant une extension spatiale	67
2.6.5	Illustration	68
2.6.6	Conclusion	69
2.7	Synthèse	70

2.1 Introduction

Si le traitement automatique du langage naturel pose depuis longtemps de gros problèmes, on peut espérer simplifier notre tâche de modélisation en restreignant notre problématique à des aspects purement spatiaux. Dans ce contexte plus réduit, des outils comme la géométrie ou la logique devraient nous permettre d'aboutir à une modélisation qui soit relativement aisée. Comme nous allons le montrer dans les deux paragraphes suivants, il n'en est rien ; nous rejoignons en cela les travaux de Nirenburg et Raskin [1987] qui démontrent que les phénomènes de représentation des raisonnements spatiaux sont complexes, à facettes multiples et parfois relativement flous.

Pour aller plus loin, il semble nécessaire de compléter la géométrie et la logique par une *composante fonctionnelle* qui se définit comme « l'ensemble des connaissances extralinguistiques de l'espace que partagent les locuteurs d'une même langue ». Nous tenterons dans un premier temps de démontrer l'importance de cette composante à l'aide des travaux de Vandeloise, qui se situent à la base de notre réflexion.

L'approche précédente étant de notre point de vue difficilement implémentable telle quelle, nous nous tournerons vers le modèle de Asher et Sablayrolles qui propose un formalisme pour les verbes de changement de lieu du français. Couplant la logique avec une approche fonctionnelle, les auteurs réalisent une étude très précise de ces verbes. Nous terminerons par le modèle de Hernández qui nous a servi de base dans l'élaboration du module de calcul des référents spatiaux que nous proposerons au cours chapitre 3. Il est à noter que le modèle de Hernández est également d'inspiration fonctionnelle.

Les approches de Vandeloise, Asher & Sablayrolles et Hernández nous semblent incontournables dans le domaine qui nous préoccupe. Elles reflètent par ailleurs la palette des différentes classes de modèles possibles en vue de rendre compte de l'usage des termes spatiaux en langue, à savoir les approches *linguistiques*, *logico-computationnelles* et *computationnelles*. Nous distinguons ces trois approches par le ou les outils qu'elles utilisent pour décrire les phénomènes mis en jeu. Tandis que la *linguistique* fournit une description détaillée de la langue en ayant recours à la langue, les modèles *computationnels* se fondent quant à eux sur les mathématiques classiques qui sont couplées à certaines heuristiques^(*). Finalement, les modèles *logico-computationnels* adjoignent aux modèles computationnels une utilisation importante de la logique.

L'influence des modèles de Vandeloise et de Hernández a été déterminante sur l'évolution de nos travaux. Les chapitres 3 et 4 préciseront les limites de ces modèles à travers un comparatif avec nos recherches.

2.2 Limites d'une approche géométrique

De par sa rigueur, la géométrie semble bien adaptée à la modélisation de l'espace. Par géométrie, nous entendons toute description de l'espace qui serait effectuée avec des outils strictement spatiaux, tels que des lignes, droites, angles et mesures.

On distingue classiquement les principales géométries suivantes : la *géométrie analytique* où l'on utilise le calcul algébrique sur les coordonnées des points, la *géométrie descriptive* où les figures sont définies par leurs projections orthogonales sur deux plans perpendiculaires, la *géométrie cotée* où l'on étudie les propriétés invariantes dans certaines transformations, la *géométrie projective* où l'on étudie la transformation des propriétés des figures par projection, enfin, la *géométrie euclidienne* à trois dimensions fondée sur le postulat d'Euclide : «Par un point extérieur à une droite, on ne peut mener qu'une seule parallèle à cette droite».

Sans viser l'une de ces géométries en particulier, les recherches actuelles s'accordent sur le fait qu'aucune d'entre elles ne peut venir à bout des nombreux problèmes attenants à la modélisation des usages spatiaux de la langue (voir en particulier [Vandeloise, 1986] et [Vieu, 1991]). Nous allons tenter d'illustrer ce fait au moyen de certains usages des prépositions “près de/loin de” et “dans”.

2.2.1 L'exemple de “près de/loin de”

Prenons le cas de la préposition “près de” dans les phrases du type “A est près de B”. Au sein de la géométrie, l'outil qui semble le plus adapté à l'étude de la préposition “près de” est la distance euclidienne classique. On l'étend aux objets comme suit : soit A et B deux objets, la distance séparant A de B est la plus courte des distances entre un point A_i appartenant à l'objet A et un point B_j appartenant à l'objet B :

$$d(A, B) = \text{Min}_{i,j} d(A_i, B_j)$$

L'usage de cette préposition paraît difficile à formaliser à l'aide de la seule distance euclidienne. Il suffit de prendre l'exemple d'un coureur cycliste qui pourrait dire “Paris est encore loin” alors qu'un aviateur serait peut-être amené à dire “Paris est très près”. De la même manière, on pourrait dire “cet électron est loin de son noyau” alors que l'on dirait “ces deux étoiles sont proches l'une de l'autre”¹.

Ces différents exemples font d'une certaine manière écho au premier chapitre. Nous avons en effet insisté au §1.4.3 sur l'importance de l'interlocuteur, c'est-à-dire celui qui perçoit le message. Ici, au contraire, c'est le locuteur qui joue un rôle essentiel. En particulier, la perspective dans laquelle il s'est placé (en tant qu'aviateur ou non aviateur). Ainsi, la conception qu'a le locuteur du monde qui l'entoure aura un rôle déterminant dans les phrases que ce dernier va énoncer.

Plus généralement, ces exemples redémontrent qu'il faut impérativement prendre en compte dans toute modélisation d'une situation de dialogue le locuteur ainsi que son

1. Exemples tirés de Vandeloise 1986.

interlocuteur. Si le locuteur et l'interlocuteur ont un rôle central à jouer dans la nature des énoncés produits, leur interaction s'inscrit de plus dans le contexte de la tâche. La tâche qui constitue l'objet de l'interaction dialogique est essentielle car c'est elle qui va conditionner pour beaucoup de contenu du dialogue. Ainsi, reprenons l'exemple de l'aviateur qui disait "Paris est très près". Imaginons à présent que le contexte de la tâche ait changé et que sa réserve de carburant soit quasi nulle, ce qui risque de le faire tomber en panne d'une seconde à l'autre. Il paraîtrait alors très naturel qu'il dise "Paris est encore loin" ou encore "Paris est trop loin".

2.2.2 L'exemple de la préposition "dans"

De la même façon, imaginons à présent que j'énonce les phrases (11) et (12) concernant un verre vide. Mon interlocuteur comprendra respectivement chacune des illustrations des figures 2.1.a et 2.1.b.

Il y a une bulle dans le verre. (11)

Il y a une mouche dans le verre. (12)

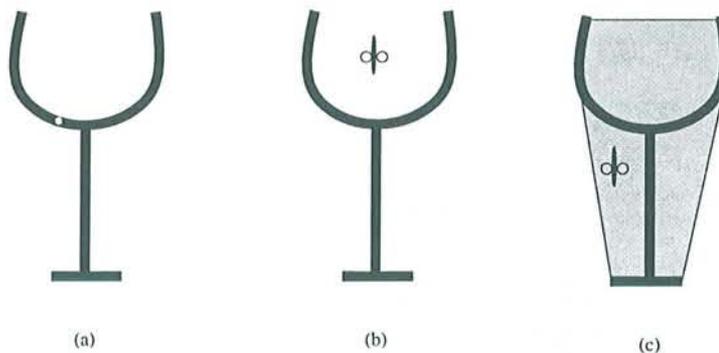


FIG. 2.1 – Un "dans" plus insaisissable qu'il n'y paraît.

On aurait alors deux types de "dans" à formaliser :

- pour la phrase (11) on pourrait utiliser : "A est dans B" si $A^* \subseteq B^*$, (A^* et B^* représentant le volume d'espace effectivement occupé par les objets en question).
- pour la phrase (12) on peut penser s'en sortir en proposant "A est dans B" si A est dans l'enveloppe convexe contenant B, mais que dire alors pour la figure 2.1.c?

Ces exemples illustrent un phénomène bien connu en linguistique qui n'est autre que la notion de contexte sur laquelle nous avons insisté au premier chapitre. Ainsi, dans l'exemple (11) et (12), on a bien l'impression que la géométrie parvient à saisir le sens premier de la préposition "dans", sens que l'on pourrait grossièrement résumer à l'aide d'une relation d'inclusion. Ce sens que les linguistes décrivent généralement à l'aide de ce qu'ils appellent la sémantique va dans un second temps s'actualiser suivant le contexte d'énonciation. Ainsi, tout ce qui est pertinent dans ce dernier va modifier (parfois radicalement) le sens premier du ou des mots considérés.

2.2.3 Conclusion

Finalement, si l'approche géométrique établit une base de travail incontournable, elle ne semble pas, seule, pouvoir fournir de description assez fine des mécanismes mis en œuvre dans les énoncés traitant de la modélisation des usages spatiaux de la langue.

Les principaux points sur lesquels elle achoppe concernent la prise en compte du locuteur et de l'interlocuteur. Nous avons également pu voir que la géométrie échouait sur les problèmes liés au contexte de la tâche. Ces remarques justifient que nous nous tournions à présent vers la prise en compte d'aspects fonctionnels afin d'obtenir de nouveaux éléments de réponse.

2.3 La prise en compte des aspects fonctionnels : une nécessité

Nous allons démontrer dans ce paragraphe qu'une bonne modélisation des usages spatiaux de la langue doit nécessairement s'appuyer sur des aspects fonctionnels. Nous illustrerons notre propos en étudiant le problème spécifique de la compositionnalité et de la transitivité en logique.

2.3.1 La compositionnalité

La logique se veut nécessairement compositionnelle, c'est-à-dire que la valeur d'une séquence d'éléments se déduit strictement d'une part de la valeur de ses éléments, d'autre part des règles de combinaison qui leur sont applicables. Si nous appliquons ce principe à des mots, nous obtenons le résultat suivant : le sens de deux termes est calculable à partir de leurs sens isolés. Néanmoins, certains énoncés expriment beaucoup plus d'information que l'on pourrait en déduire *a priori* d'une simple combinaison de chacun des termes pris isolément. Ce fait général apparaît également au niveau des énoncés spatiaux. Ainsi, prenons le cas de la phrase (13) qui implique plus que la simple proximité des deux éléments considérés, elle indique en plus que Paul se trouve vraisemblablement assis à son bureau et qu'il y travaille.

Paul est à son bureau. (13)

Dans ces phrases du type "*X est à Y*", il y a une activité particulière de l'entité localisée par rapport à l'entité localisatrice que seuls des aspects fonctionnels sont à même de prendre en compte. Vandeloise [1988] propose de décrire les phrases du type "*X est à Y*" en ayant recours à l'existence ou l'inexistence d'une propriété d'interaction typique ou activité particulière de la cible^(*) par rapport au site^(*) qu'il dénomme "routine".

Tout comme la compositionnalité, la transitivité peut poser quelques problèmes en modélisation du langage naturel. Les aspects fonctionnels ont là encore un rôle décisif à jouer comme cela apparaîtra dans le paragraphe suivant.

2.3.2 La transitivité

La transitivité n'est pas toujours respectée² pour la préposition “*sur*” comme le démontrent les phrases (14) et (15) dont on ne peut pas déduire la phrase (16). Il en va de même pour la préposition “*dans*” qui semble pourtant transitive par essence, voir les phrases (17), (18) et (19).

Le livre est sur la table. (14)

La table est sur le sol. (15)

* *Le livre est sur le sol.* (16)³

L'abeille est dans la rose. (17)

La rose est dans le vase. (18)

* *L'abeille est dans le vase.* (19)

Vieu et Aurnague démontrent que les aspects fonctionnels sont indispensables pour saisir précisément le sens associé aux prépositions “*dans*” [Vieu, 1991] et “*sur*” [Aurnague, 1991]. D'autre part, les aspects fonctionnels permettent de calculer les éventuelles transitivités pour les phrases faisant appel à ces prépositions. On pourra se reporter à Aurnague et Vieu [1993] pour une très bonne synthèse globale de ces travaux.

2.3.3 Conclusion

Ces différents points montrent qu'en préalable à une formalisation trop brutale de l'espace il est nécessaire de préciser le fonctionnement précis des opérateurs spatiaux du langage naturel. En particulier, la géométrie peut être qualifiée d'outil “sans âme” dans le sens où elle ne parvient pas à saisir toute la matière linguistique. Une modélisation fine des aspects spatiaux passe au contraire par une nécessaire prise en compte d'aspects fonctionnels comme nous avons pu le démontrer pour le problème de la compositionnalité et de la transitivité au sein de la logique.

Ce paragraphe et le précédent justifient donc l'importance d'une étude linguistique fine qui tienne compte de tous les niveaux d'analyse, dont la pragmatique. L'analyse de Vandeloise semble être en mesure d'apporter quelques éléments de réponse, comme nous allons le voir tout de suite.

2.4 Le modèle de Vandeloise

L'approche de Vandeloise [1986] fournit une description fonctionnelle des prépositions spatiales françaises. Le cadre théorique de cette approche se fonde sur la notion de ressemblance de famille introduite par Wittgenstein [1953]. Après avoir présenté cette notion ainsi que celle d'impulsion, nous en viendrons aux bases de l'approche de Vandeloise qui se

2. À l'inverse de la compositionnalité, la transitivité constitue une propriété qui n'est pas toujours souhaitable.

3. Les phrases précédées d'un astérisque “*” désignent des énoncés incorrects tandis que celles précédées d'un point d'interrogation “?” attestent d'un énoncé douteux mais acceptable.

fonde sur la notion de cible/site ainsi que sur la prise en compte des aspects fonctionnels du monde qui nous entoure. Au terme de cette étude, nous verrons une illustration de l'ensemble de ces principes à l'aide de l'usage des prépositions "sur/sous".

2.4.1 Prolégomènes: les notions de ressemblance de famille et d'impulsion

Les travaux de Vandeloise se fondent sur la notion centrale de ressemblance de famille qu'il définit comme suit :

Ressemblance de famille: concept représenté par la combinaison des traits qui le caractérisent, sachant que chaque trait n'est pas toujours nécessaire ni suffisant, il s'agit d'une structuration qui permet aux membres d'une catégorie d'être reliés les uns aux autres, sans avoir une propriété en commun qui définisse la catégorie.

Ce qui réunit ces membres, ce sont des similarités, des ressemblances qui s'entrecroisent, se recouvrent partiellement. On trouvera deux illustrations de cette notion d'une part à la figure 2.2.(a) qui représente quelques exemplaires de la catégorie "oiseau", d'autre part, à la figure 2.2.(b)⁴ qui vérifie bien la propriété que chaque exemplaire de la catégorie partage au moins une propriété avec un autre membre de la catégorie.

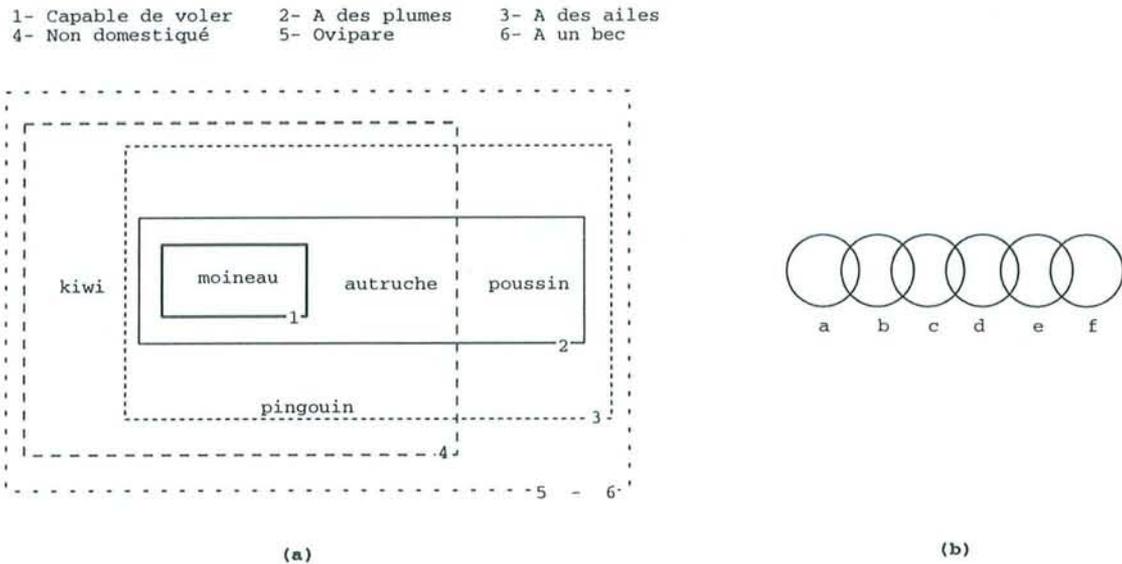


FIG. 2.2 – La notion de ressemblance de famille.

La notion de ressemblance de famille, qui a été déterminante dans l'évolution de la théorie de Rosch concernant la catégorisation⁵, permet à Vandeloise de caractériser

4. À titre d'illustration, "a", "b", "c", "d", "e" et "f" peuvent par exemple désigner les membres d'une famille. L'individu "a" présentant par exemple la même couleur de cheveux que l'individu "b", ce dernier présentant à son tour une ou plusieurs similitudes avec l'individu "c", etc.

5. Voir en particulier [Kleiber, 1990a] pour un exposé détaillé du passage de la "théorie standard" à la "théorie étendue" de la catégorisation, cette dernière se fondant essentiellement sur la notion de ressemblance de famille.

chaque préposition à l'aide d'une règle ou plusieurs dont chaque trait n'est ni nécessaire, ni suffisant. Vandeloise cherche dans la mesure du possible à caractériser chaque préposition par une règle d'usage unique, mais, pour certaines prépositions (par exemple "*devant/derrière*") plusieurs règles d'usage doivent être introduites. Ces règles couvrent l'ensemble des usages, mais deux cas sont possibles : soit les règles sont en distribution complémentaire, soit leurs distributions se chevauchent, le premier cas étant préférable bien entendu. Ce dernier point nous écarte sensiblement d'une approche en termes de conditions nécessaires et suffisantes⁶.

Il peut sembler difficile d'expliquer d'un point de vue synchronique⁷ comment les distributions complexes de certaines expressions spatiales (représentées par les règles d'usage) peuvent être reliées à un signifiant⁸ simple. Afin d'expliquer l'acquisition d'un réseau si complexe de significations, Vandeloise suppose l'existence d'une règle d'usage privilégiée (dénommée impulsion) qui constituerait en réalité le nœud d'accès à l'ensemble de ce réseau. Le développement des autres règles d'usages se ferait donc à partir de la dite impulsion. Vandeloise définit le concept hypothétique d'impulsion comme suit :

Impulsion : «*il a existé à l'origine des prépositions spatiales une correspondance transparente et univoque entre leur signifiant et leur signifié*».

2.4.2 La prise en compte de la pragmatique chez Vandeloise

Vandeloise fait reposer sa description de la sémantique des prépositions spatiales sur un certain nombre de concepts fondamentaux que nous allons décrire en détail.

1. La notion de *cible, site*⁹. Se fondant sur le principe extra-linguistique suivant : «*Un objet dont la position est incertaine ne peut être localisé sans référence à une entité dont la position est mieux connue*», Vandeloise définit la notion de *cible/site*. Tandis que la cible joue le rôle d'entité localisée, le site joue celui d'entité localisatrice.

Quelles sont les caractéristiques de la cible et du site? L'auteur mentionne que la position de la cible est généralement une position nouvelle, tandis que celle du site est souvent plus ancienne. De plus, alors que la cible est petite ou difficile à repérer, le site est généralement massif et facile à distinguer (saillant). Enfin, la cible est souvent mobile et susceptible de bouger, alors que le site est généralement immobile et stable.

6. Le terme de conditions nécessaires et suffisantes a été explicité au paragraphe 1.4.4

7. Tandis que la *diachronie* fournit une description de l'évolution d'une ou plusieurs langues, la *synchronie* se préoccupe du même problème mais en un point donné du temps.

8. Tandis que le *signifiant* désigne la forme extérieure c'est-à-dire l'ensemble des sons ou des lettres, le *signifié* représente le sens. L'union d'un *signifiant* et d'un *signifié* représente ce que l'on appelle un *signe*. Pour fixer les idées, prenons l'exemple du mot "*cheval*", qui est un *signe*, il se constitue d'un *signifiant*, correspondant à une suite de sons à l'oral ou une suite de lettres à l'écrit, et d'un *signifié* qui représente le sens tel que les dictionnaires tentent de le définir (pour de plus amples détails, voir [Saussure, 1968]).

9. Ces notions ont été introduites, en d'autres termes, avant Vandeloise par Miller et Johnson-Laird, mais également par Talmy et Herskovits.

Vandeloise observe que les prépositions spatiales ne sont pas toujours *converses*¹⁰ comme le montrent les phrases (20) et (21), il explique ce phénomène à l'aide de la symétrie qui peut éventuellement exister entre la cible et le site et énonce la règle suivante : «*Si le rôle de la cible et du site peuvent être échangés, les prépositions associées sont converses.*» Les phrases (22) et (23) illustrent ce propos pour les prépositions “*devant/derrière*” :

La pierre est devant la maison. (20)

* *La maison est derrière la pierre.* (21)

Le chêne est devant le peuplier. (22)

Le peuplier est derrière le chêne. (23)

2. Idéalisation des arguments. La cible et le site qui interviennent dans les relations spatiales sont souvent idéalisés au niveau de leur dimensionalité (dimension 0, 1, 2 ou 3) : par exemple, un “*étang*” est généralement idéalisé par sa surface (un “*plan d'eau*”), on pourra alors dire d'un canard qu'il nage “*sur*” l'eau et plonge “*sous*” l'eau plutôt que “*dans*” l'eau.

L'idéalisation peut également porter sur le fait qu'à un instant donné du discours, le locuteur ne conçoit qu'une portion de l'entité qu'il mentionne. Pour illustrer ce point, il suffit de prendre l'exemple d'un curé dont on dirait qu'il se trouve “*au milieu de l'église*”. Dans ce cas, seul le sol de l'église est à considérer, le curé n'étant pas situé en l'air au milieu du volume que délimite l'église.

3. Le principe de fixation. Vandeloise définit ce principe comme suit :

Principe de fixation : «*un objet peut être qualifié par rapport à sa position usuelle, même si sa position diffère au moment de l'énonciation.*»

Pour illustrer ce principe, il suffit de prendre les prépositions “*le dessus de*” et “*le dessous de*” qui sont généralement déduites à partir de la verticale gravitationnelle, cependant, dans le cas d'une bouteille, le dessous de cette dernière ne change pas lorsqu'on la renverse, “*le dessus de*” et “*le dessous de*” sont pour ainsi dire fixés par la position normale de la bouteille. On parlera alors d'*orientation intrinsèque* pour ces objets.

4. Le principe de transfert. Voici la définition de ce principe :

Principe de transfert : «*le locuteur a la faculté de se déplacer en tout point utile à la perspective selon laquelle il conçoit la scène objective qu'il décrit.*»

Le site, couplé à un point de perspective donné, va nous permettre d'obtenir une direction qui servira dans le choix entre les prépositions “*devant, derrière, à droite, à gauche, dessus, dessous*”. Le point de perspective pourra être soit explicite (on parlera alors d'*orientation extrinsèque*) soit implicite (on parlera alors d'*orientation déictique*), dans ce dernier cas le point de perspective sera souvent donné par la position du locuteur.

10. Deux prépositions sont converses si l'une est l'inverse de l'autre. Par exemple “*à droite de*” et “*à gauche de*” sont converses.

Vandeloise mentionne le fait que si les positions du locuteur sont en nombre illimité, la conversation, elle, ne fonctionne de manière satisfaisante qu'aussi longtemps que l'interlocuteur est capable de suivre la trajectoire mentale du premier. Cette notion de connaissance partagée sera reprise par la notion de cadres du chapitre 4 où une rupture de cadre entre deux locuteurs sera généralement source d'ambiguïté. Notons enfin que le concept de trajectoire mentale partagée s'inscrit en fait dans le contexte plus général de la pragmatique dans le sens où le discours a un rôle essentiel à jouer dans l'interprétation d'un énoncé. Ce point fait écho au chapitre 1 qui insistait sur l'importance de la notion de contexte partagé.

2.4.3 Les aspects fonctionnels

Vandeloise répartit les caractéristiques fonctionnelles des relations spatiales en cinq groupes de "traits universels", ces traits étant considérés comme des primitives, donc non définissables dans la théorie de l'auteur :

1. Anthropomorphie et forme du corps humain. Si comparé à d'autres langues¹¹, le français ne fait qu'un usage très pauvre du lexique associé aux parties du corps humain pour décrire l'espace, la symétrie et les fonctions de ce dernier déterminent deux directions privilégiées : la direction frontale et la direction latérale (figure 2.3). L'anthropomorphie nous conduit à orienter certains objets relativement à nous même : pour un lit par exemple, on peut parler de la tête du lit, dans le cas d'une armoire, de la face de l'armoire.

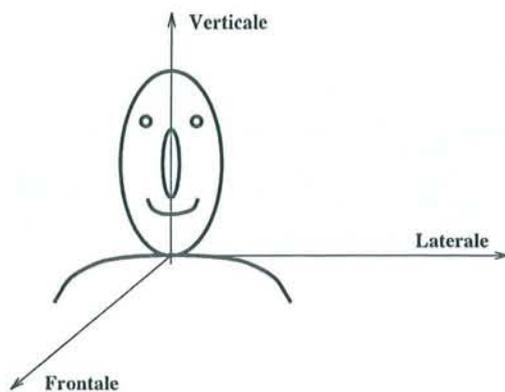


FIG. 2.3 – Les trois directions de base dans le repère du corps humain.

De même, la position selon laquelle deux interlocuteurs se parlent, face à face, à une courte distance l'un de l'autre (appelée l'interaction canonique) a une influence directe sur les prépositions "*devant/derrière*". On distingue alors deux orientations contextuelles que peut donner un interlocuteur à un objet qu'il voit, selon qu'il oriente ce dernier en "miroir" (figure 2.4.(a)) ou en "tandem" (figure 2.4.(b)). Ainsi, tandis que la première orientation s'applique à des objets non intrinsèquement orientés, la seconde s'applique généralement à des objets en mouvement.

11. Vandeloise cite les indiens Tarascos et Coras.

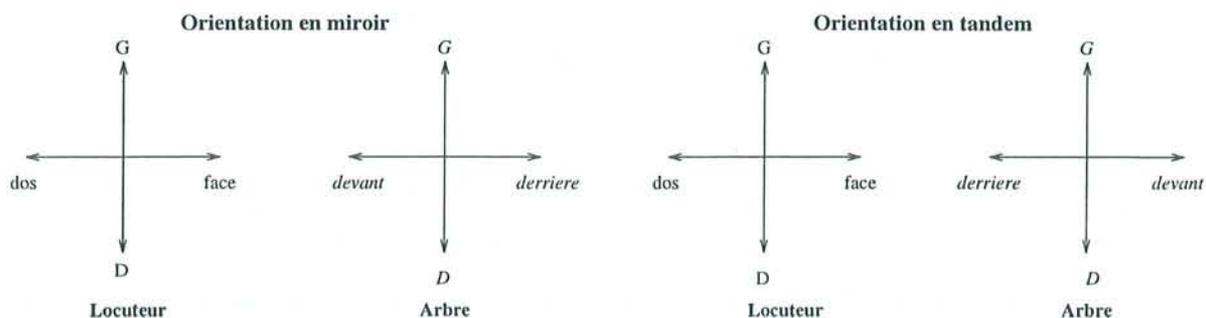


FIG. 2.4 – Les orientations en miroir et en tandem.

2. La “physique naïve”. Introduite par Hayes [1986], cette partie de l’intelligence artificielle consiste à formaliser l’espace de sens commun, ainsi que l’ensemble des propriétés du monde physique, c’est-à-dire les connaissances que nous avons du monde physique réel qui nous entoure. Vandeloise s’attache aux croyances portant sur le monde qui sont partagées par tous les usagers d’une même langue comme le fait, par exemple, qu’un objet qui porte un autre objet est généralement plus grand que celui-ci, plus proche de la terre, partiellement caché par lui, etc. Vandeloise conclut en soulignant que la physique naïve diffère autant de la physique scientifique que les langues naturelles des langages formalisés.

3. L’accès à la perception. Elle permet de rendre compte de l’usage des prépositions “derrière” et “sous” par l’inaccessibilité de la cible relativement au site. Ainsi, dans les phrases (24) et (25), illustrées respectivement aux figures 2.5.(a) et 2.5.(b), l’usage de la préposition “derrière” s’explique par le fait que dans les deux cas, la cible n’est pas perceptible: dans le premier cas parce que la cible (Paul), se trouvant dans le dos de Jean, est forcément hors du champ visuel de ce dernier, et dans le second cas parce que le peuplier empêche Jean de voir Paul.

Paul est derrière Jean. (24)

Paul est derrière le peuplier. (25)

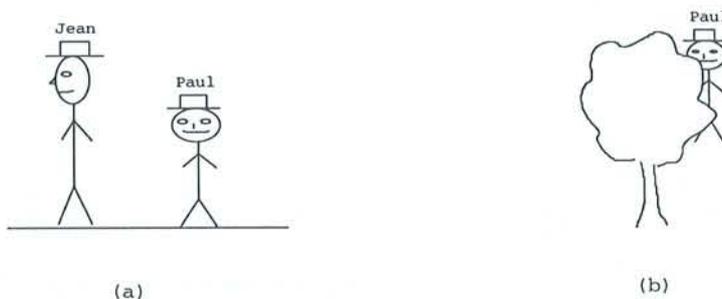


FIG. 2.5 – Deux illustrations de la préposition “derrière”.

4. La rencontre potentielle. Elle permet de rendre compte de l’usage des prépositions “avant/après” suivant une sorte de chemin virtuel qui relie la cible, le site à un

élément non exprimé mais défini par le contexte: le *pôle*. Vandeloise distingue trois cas selon que la cible, le site sont statiques ou non :

- la cible et le site sont statiques,
- la cible ou le site est statique,
- la cible et le site sont mobiles.

Ainsi dans une phrase du type “*le chêne est avant le peuplier*”, où la cible et le site sont statiques, c’est le chemin du locuteur (qui joue ici le rôle de *pôle*) vers les arbres qui est considéré: le locuteur rencontrera (réellement ou virtuellement) le chêne avant de rencontrer le peuplier. Ce chemin virtuel peut être déterminé par le regard ou suivre au contraire un tracé plus complexe comme celui d’un itinéraire sur une carte géographique par exemple.

5. L’orientation générale et l’orientation latérale. La première orientation peut être définie comme une ressemblance de famille dont les principaux traits sont : *la direction frontale, la direction du mouvement, la direction du regard, la direction des autres organes de perception: le goût, l’ouïe, etc..*

L’orientation latérale est également une ressemblance de famille dont les principaux traits sont la direction latérale et la perpendiculaire à l’orientation générale. Ces notions sont utilisées pour caractériser les prépositions “*à droite/à gauche*” et “*devant/derrière*” (pour un exposé détaillé, voir [Vandeloise, 1986]).

2.4.4 Un exemple: les prépositions “*sur/sous*”

Vandeloise énonce la règle d’usage **S** suivante pour les prépositions “*sur/sous*” :

S: A est “*sur/sous*” B si sa cible est le deuxième/le premier élément de la relation porteur/porté et son site, le premier/le deuxième élément de cette relation.

La relation porteur/porté se comporte comme une ressemblance de famille dont les différents traits sont les suivants :

- **Caractéristique A** (*ordre sur l’axe vertical*): si A est “*sur/sous*” B, la cible est généralement plus haute/plus basse que le site.

Le “généralement” est précisé pour des exemples du type: “*la mouche est sur le plafond*”.

- **Caractéristique B** (*contact*): si A est “*sur*” B, il y a généralement un contact direct entre la cible et le site.

Vandeloise illustre cette règle à l’aide de deux exemples: le premier illustre la possibilité que le contact ne soit pas nécessairement direct comme cela apparaît à la phrase (26), en imaginant que trois livres A, B et C se trouvent empilés sur une table (le livre A se trouvant alors au sommet de la pile). Le second exemple démontre que tous les contacts indirects ne sont pas permis, on peut illustrer ce fait en prenant le cas d’une casserole (possédant un couvercle) qui serait posée sur une table et pour laquelle la phrase (27)

serait inappropriée :

Le livre A est sur la table. (26)

* *Le couvercle est sur la table.* (27)

- **Caractéristique C** (*accès à la perception*) : si A est “*sous*” B, la cible est généralement rendue inaccessible à la perception par le site.

Vandeloise mentionne le fait que cette accessibilité diminue avec la capacité du site à couvrir la cible.

- **Caractéristique D** (*dimension de la cible et du site*) : dans les relations A est “*sur/sous*” B, la cible est généralement plus petite que le site.

- **Caractéristique E** (*opposition à la pesanteur*) : si A est “*sur*” B, l’action du site s’oppose à l’action de la pesanteur sur la cible.

Cette caractéristique autorise l’utilisation de la préposition “*sur*” dans la phrase (28) pour laquelle aucune des autres caractéristiques exposées ci-dessus ne peut expliquer l’usage.

Les feuilles sont sur l’arbre. (28)

Remarques :

- La relation porteur/porté est bien une ressemblance de famille dans le sens où aucune des caractéristiques exposées ci-dessus n’est ni nécessaire (prendre l’exemple d’une mouche marchant sur le plafond qui violerait la caractéristique A de l’ordre sur la verticale), ni suffisante (voir la phrase (29) illustrée à la figure 2.6 pour laquelle la caractéristique A n’est pas suffisante). Ce point semble cependant nuancable. Ainsi, si l’on se place dans la perspective de la mouche en utilisant par exemple le principe de transfert, on serait alors plus à même de dire : “*la mouche marche sur le plafond.*”

**La lampe est sur la table.* (29)

- L’application de la relation porteur/porté est souvent trop lâche (voir l’exemple de la figure 2.7 pour laquelle on préférerait la phrase (31) à la phrase (30)).

**Le balai est sur le mur.* (30)

Le balai est contre le mur. (31)

Pour interdire certains usages de la préposition “*sur*” au profit des prépositions “*à*” et “*contre*”¹² Vandeloise propose de prendre en considération l’aspect actif, intermédiaire ou

12. L’auteur introduit le concept de “restrictions de sélection”.

FIG. 2.6 - * *La lampe est sur la table.*FIG. 2.7 - * *Le balai est sur le mur.*

passif de la relation porteur/porté qu'il suggère de compléter par des notions d'interaction (verticale ou horizontale) entre la cible et le site et d'interdépendance de leur position (voir [Vandeloise, 1986] p. 200 à 205 pour un exposé détaillé et [Vandeloise, 1986] p. 205 à 207 pour les restrictions de sélection pesant sur la préposition "*sous*" au profit de la préposition "*derrière*").

2.4.5 Conclusion

On notera tout d'abord l'importance de la notion de cible et de site qui possèdent une validité du point de vue de la psychologie cognitive comme cela apparaît dans les travaux de Miller et Johnson-Laird [1976]. Par ailleurs, leurs études confirment que les êtres humains conceptualisent un espace relatif dans lequel les entités sont repérées les unes par rapport aux autres plutôt qu'à l'aide d'un espace absolu qui utiliserait, pour localiser les objets, un système de coordonnées.

De cette approche, outre son caractère novateur et original, nous retiendrons surtout l'apport indéniable des aspects fonctionnels qu'il est nécessaire de mettre en œuvre dans tout système de modélisation spatiale digne de ce nom. Enfin, l'un des points forts de ces travaux réside dans la clarté avec laquelle les outils de base et la méthodologie sont définis. Vandeloise a réussi à éviter l'écueil qui aurait consisté à intégrer l'ensemble des concepts fonctionnels de manière désordonnée, ce qui donne à son approche une structure cohérente.

La notion de fonctionnalité sera omniprésente au sein de la notion de cadre. Les chapitres 3 et 4 nous offriront l'occasion de mentionner quelques limites de l'approche de Vandeloise. Par ailleurs, le reproche central que l'on peut faire à l'étude de Vandeloise provient de son caractère peu implémentable dans une perspective informatique. Ce fait tient essentiellement à la notion de ressemblance de famille qui est sous-jacente au modèle proposé. La notion de ressemblance de famille semble en effet difficilement

programmable directement de par ses spécificités où aucune caractéristique la composant n'est ni nécessaire, ni suffisante.

Ce dernier point justifie notre intérêt pour le modèle formel de Asher et Sablayrolles que nous allons décrire à présent.

2.5 Le modèle de Asher et Sablayrolles

2.5.1 Introduction

En regard des travaux de Vandeloise, les recherches que nous allons présenter se placent d'emblée dans une perspective d'implémentation informatique. Ainsi, l'approche de Asher et Sablayrolles [1995] réalise une étude précise de la sémantique spatio-temporelle des verbes de changement de lieu (intransitifs) et des prépositions spatiales¹³.

2.5.2 Taxinomie des verbes de mouvement

Asher et Sablayrolles distinguent quatre classes de verbes de mouvement :

- *Les verbes de changement de lieu*: ces verbes, comme par exemple “partir”, “rentrer” ou “éloigner”, imposent que la cible¹⁴ change de localisation par rapport au site, dans (32), “Jean” se trouve à l'extérieur de la “maison” au début du processus et à l'intérieur à la fin.

Jean rentre dans la maison. (32)

- *Les verbes de changement d'emplacement*: cette classe comprend des verbes comme “déplacer, circuler, descendre”. Les verbes de changement d'emplacement imposent seulement que la cible change de localisation par rapport à une partie du site. Dans la phrase (33), “Jean” reste à l'intérieur de “la ville” durant tout le processus, même si sa position doit obligatoirement changer par rapport à un point particulier de “la ville”.

Jean se déplace dans la ville. (33)

- *Les verbes de changement potentiel d'emplacement*: ces verbes permettent que la cible change de localisation par rapport à une partie du site mais ils ne l'imposent pas. Cette classe comprend des verbes comme “courir, voler, danser”. Par exemple, dans (34), “Jean” ne change pas du tout de localisation.

13. Ce paragraphe se fonde en partie sur Hathout [1994].

14. Les notions de cible et de site sont celles définies précisément auparavant au sein du modèle de Vandeloise.

Jean court sur place. (34)

Le fait de pouvoir se construire avec l'expression “*sur place*” est un caractère propre aux verbes de changement potentiel d'emplacement, comme en témoignent les contre-exemples suivants :

* *Jean sort sur place.* (35)

* *Jean se déplace sur place.* (36)

- *Les verbes de changement de posture*: ces verbes comme “*se pencher, se baisser, s'agenouiller*”, interdisent à la cible de changer d'emplacement ou de lieu.

Asher et Sablayrolles décrivent ensuite précisément les changements de classe des verbes de mouvement : un syntagme^(*) verbal construit à partir d'un verbe de mouvement et d'un syntagme prépositionnel peut ne pas appartenir à la même classe que le verbe employé seul.

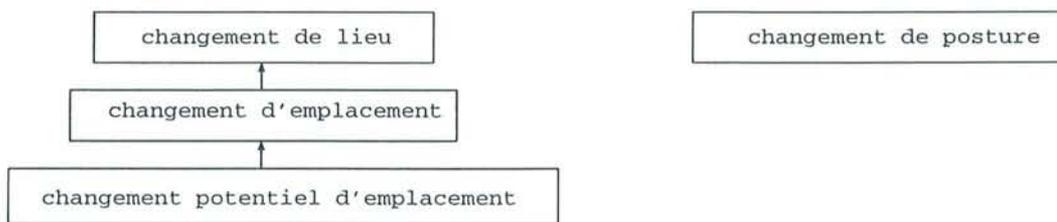


FIG. 2.8 – *Changements de classe des verbes de mouvement.*

Un verbe de changement potentiel d'emplacement peut être tête d'un SV¹⁵ qui décrit un changement d'emplacement (37) ou un changement de lieu (38). De même, un verbe de changement d'emplacement peut donner un SV de changement de lieu. Les verbes de changement de lieu et de changement de posture ne peuvent, en revanche, être tête que d'un SV de leur classe. La figure 2.8 résume schématiquement les changements possibles de classe des SV de mouvement.

Jean court partout dans sa cuisine. (37)

Jean court jusqu'à la cuisine. (38)

Dans le reste de ce paragraphe, nous ne considérerons que les verbes de changement de lieu car ce sont les seuls pour lesquels Asher et Sablayrolles ont fourni une sémantique spatio-temporelle.

15. SV : syntagme verbal.

2.5.3 Sémantique des verbes de changement de lieu

La description de Asher et Sablayrolles repose sur un découpage de l'espace en sept zones (figure 2.9) définies relativement au site et qui sont des fonctions du site, de la cible et du contexte [Aurnague *et al.*, 1993]. On peut définir à l'aide de ces zones la sémantique spatio-temporelle d'un verbe comme "partir" de la manière suivante¹⁶ :

$$\begin{aligned}
 \text{partir}(e) \rightarrow & \\
 & P(\text{source}(e), \text{ZHI}(\text{cible}(e), \text{Lref}(e))) \wedge \\
 & P(\text{chemin}(e), \text{ZHE}(\text{cible}(e), \text{Lref}(e))) \wedge \\
 & P(\text{but}(e), \text{ZEE}(\text{cible}(e), \text{Lref}(e))) \wedge \\
 & \text{Initial}(e, \text{Lref}(e))
 \end{aligned}
 \tag{39}$$

Cette formule décrit que e est un événement (le départ) dont la localisation se fait par rapport à $\text{Lref}(e)$, à savoir le site considéré dans la relation spatiale introduite par le verbe. Les trois premiers termes dans la conséquence établissent que le début, le milieu et la fin de e doivent se dérouler respectivement à l'intérieur du site, à proximité (à l'extérieur) du site et loin du site. Le quatrième terme conjoint indique que e est de polarité initiale par rapport au site. Les auteurs s'appuient ici sur le formalisme développé par Laur [1993] qui est très similaire à celui de Boers [1987] en répartissant les verbes de changement de localisation en verbes de polarité de type initiale, médiane et finale :

- polarité initiale : s'éloigner, partir, décoller, etc.
- polarité médiane : passer, etc.
- polarité finale : s'approcher, arriver, entrer, etc.

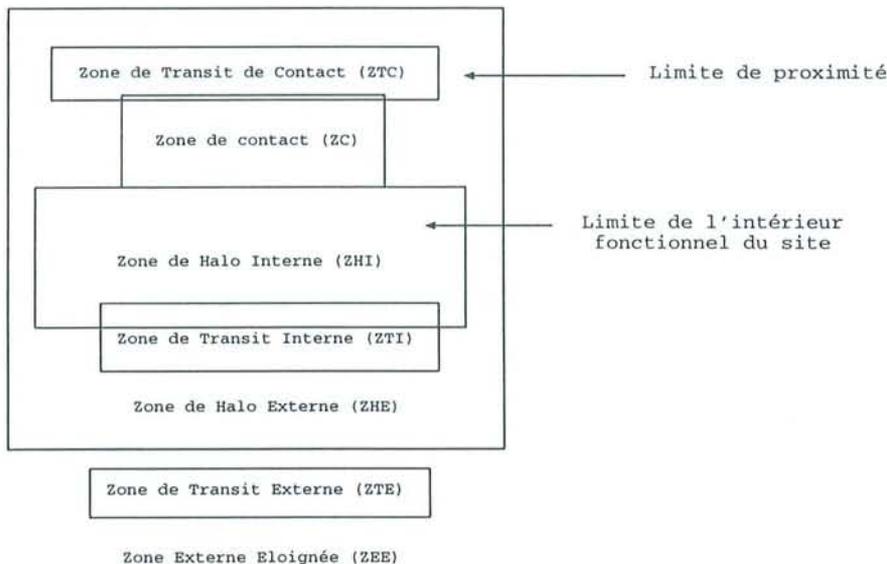


FIG. 2.9 – Les sept zones de localisation d'après Asher et Sablayrolles 1995.

Dans le cas du verbe "rentrer", l'événement est une entrée ; son début, son milieu et sa fin ont lieu respectivement à proximité (à l'extérieur) du site, à la limite du site et à

16. ZHI : Zone de Halo Interne, ZHE : Zone de Halo Externe, etc.

l'intérieur du site ; sa polarité est finale par rapport au site :

$$\begin{aligned}
 \text{rentrer}(e) \rightarrow & \\
 & P(\text{source}(e), ZHE(\text{cible}(e), Lref(e))) \wedge \\
 & P(\text{chemin}(e), ZTI(\text{cible}(e), Lref(e))) \wedge \\
 & P(\text{but}(e), ZHI(\text{cible}(e), Lref(e))) \wedge \\
 & \text{Final}(e, Lref(e))
 \end{aligned} \tag{40}$$

Finalement, on obtient 10 groupes de verbes selon les zones de localisation qu'ils mettent en jeu, ce qui permet de donner une sémantique spatio-temporelle à chacun des verbes de changement de lieu (intransitifs).

Le cadre formel utilisé pour décrire la sémantique spatio-temporelle des verbes de changement de lieu permet de décrire celle des prépositions spatiales à l'aide des mêmes notions que pour les verbes. Par exemple, (41) définit la sémantique spatio-temporelle de “pour” et (42) définit celle de la préposition “de” dans le cas où elle a le sens de “l'intérieur de”, “dedans”.

$$\begin{aligned}
 \text{pour}(\text{site}, e) \rightarrow & \\
 & P(\text{but}(e), ZEE(\text{cible}(e), \text{site})) \wedge \\
 & \text{Final}(e, \text{site})
 \end{aligned} \tag{41}$$

$$\begin{aligned}
 \text{de}(\text{site}, e) \rightarrow & \\
 & P(\text{source}(e), ZHI(\text{cible}(e), \text{site})) \wedge \\
 & \text{Initial}(e, \text{site})
 \end{aligned} \tag{42}$$

Ces formules donnent, pour les groupes prépositionnels “de Paris” et “pour Nancy” les représentations sémantiques (43) et (44) pour lesquels $STref(x)$ représente une nouvelle entité du discours désignant la trajectoire parcourue par x tout au long de sa “vie”.

$$\begin{aligned}
 \exists x, \text{Paris}(x) \wedge & \\
 & P(\text{source}(e), ZHI(\text{cible}(e), STref(x))) \wedge \\
 & \text{Initial}(e, STref(x))
 \end{aligned} \tag{43}$$

$$\begin{aligned}
 \exists y, \text{Nancy}(y) \wedge & \\
 & P(\text{but}(e), ZEE(\text{cible}(e), STref(y))) \wedge \\
 & \text{Final}(e, STref(y))
 \end{aligned} \tag{44}$$

La sémantique proposée par Asher et Sablayrolles permet de construire, compositionnellement des représentations pour ce qu'ils appellent des *complexes verbaux*, à savoir des syntagmes composés d'un verbe de déplacement et d'un ou plusieurs syntagmes prépositionnels. Le calcul de ces représentations est fait à l'aide de huit règles ; ces règles ont comme principaux paramètres la nature directionnelle *vs.* positionnelle de la préposition et l'identité de la polarité éventuelle de la préposition avec celle du verbe. Les prépositions positionnelles n'ont pas de polarité. Dans le cas de la composition de ces prépositions avec un verbe de déplacement, seule la polarité du verbe est prise en considération pour le calcul de la représentation sémantique du SV. Nous ne présenterons ici que les règles qui concernent les prépositions directionnelles (45) à (48) ; les règles relatives aux propositions positionnelles mettent en œuvre des formes de composition

identiques à celles qui sont discutées ci-dessus.

$$\begin{aligned} & \text{directionnelles}(P) \wedge \text{polarite}(P) = \text{polarite}(v) \\ \rightarrow & \quad \text{Lref}(e) = \text{ZonePrep}(\text{cible}(e), \text{STref}(N)) \end{aligned} \quad (45)$$

$$\begin{aligned} & \text{directionnelle}(P) \wedge \text{polarite}(P) \neq \text{polarite}(v) \wedge \text{polarite}(v) = I \\ \rightarrow & \quad P(\text{source}(e), \text{ZonePrep}(\text{cible}(e), \text{STref}(N))) \wedge \\ & \text{Lref}(e) = ? \end{aligned} \quad (46)$$

$$\begin{aligned} & \text{directionnelle}(P) \wedge \text{polarite}(P) \neq \text{polarite}(v) \wedge \text{polarite}(v) = M \\ \rightarrow & \quad P(\text{chemin}(e), \text{ZonePrep}(\text{cible}(e), \text{STref}(N))) \wedge \\ & \text{Lref}(e) = ? \end{aligned} \quad (47)$$

$$\begin{aligned} & \text{directionnelle}(P) \wedge \text{polarite}(P) \neq \text{polarite}(v) \wedge \text{polarite}(v) = F \\ \rightarrow & \quad P(\text{but}(e), \text{ZonePrep}(\text{cible}(e), \text{STref}(N))) \wedge \text{Lref}(e) = ? \end{aligned} \quad (48)$$

Dans ces formules, $\text{ZonePrep}(\text{cible}(e), \text{STref}(N))$ est la zone définie par le groupe prépositionnel et N est le référent de discours associé au SN¹⁷ complément de la préposition. L'utilisation de ces règles peut être illustrée en détaillant le calcul de la représentation sémantique du SV “partir de Paris pour Nancy”. Ce calcul se déroule en deux étapes : on commence par composer la représentation de “partir” donnée en (39) avec celle de “de Paris”, donnée en (43). La préposition “de” possède dans ce cas une utilisation directionnelle et elle a la même polarité que “partir” ; on doit donc utiliser la règle (45) pour composer les représentations de ces deux constituants. On obtient ainsi l'équation :

$$\exists x, \text{Paris}(x) \wedge \text{Lref}(e) = \text{ZHI}(\text{cible}(e), \text{STref}(x)) \quad (49)$$

qui reportée dans (39), donne :

$$\begin{aligned} \exists x, & \text{Paris}(x) \wedge \\ & P(\text{source}(e), \text{ZHI}(\text{cible}(e), \text{ZHI}(\text{cible}(e), \text{STref}(x)))) \wedge \\ & P(\text{chemin}(e), \text{ZHE}(\text{cible}(e), \text{ZHI}(\text{cible}(e), \text{STref}(x)))) \wedge \\ & P(\text{but}(e), \text{ZEE}(\text{cible}(e), \text{ZHI}(\text{cible}(e), \text{STref}(x)))) \wedge \\ & \text{Initial}(e, \text{ZHI}(\text{cible}(e), \text{STref}(x))) \end{aligned} \quad (50)$$

Notons que, dans tous les cas, la composition préserve les caractéristiques du verbe comme par exemple sa polarité ; en d'autres termes, elle est compatible avec la définition syntaxique de ce dernier comme tête du syntagme verbal. La représentation du syntagme verbal complet (52) est calculée lors de la seconde étape en composant (50) avec (44). Le complexe verbal (50) étant de polarité initial et le SP¹⁸ étant de polarité finale, on utilise (46) qui produit l'ajout de la condition :

$$\exists y, \text{Nancy}(y) \wedge P(\text{but}(e), \text{ZEE}(\text{cible}(e), \text{STref}(y))) \quad (51)$$

17. SN : syntagme nominal.

18. SP : syntagme prépositionnel.

à la représentation de “partir de Paris”. On obtient donc :

$$\begin{aligned}
& \exists x, \exists y, Paris(x) \wedge Nancy(y) \wedge \\
& P(source(e), ZHI(cible(e), ZHI(cible(e), STref(x)))) \wedge \\
& P(chemin(e), ZHE(cible(e), ZHI(cible(e), STref(x)))) \wedge \\
& P(but(e), ZEE(cible(e), ZHI(cible(e), STref(x)))) \wedge \\
& P(but(e), ZEE(cible(e), STref(y))) \wedge \\
& Initial(e, ZHI(cible(e), STref(x)))
\end{aligned} \tag{52}$$

Cet exemple illustre le fait que les SP spatiaux ne remplissent pas tous le même rôle (sémantique) par rapport au verbe dont ils dépendent. On a ainsi deux modes de composition sémantique des verbes et des complexes verbaux avec les syntagmes prépositionnels. Le premier, illustré par (50), est l’instanciation du lieu de référence : le SP définit le site de la relation spatiale introduite par le verbe. En d’autres termes, le SP complète la représentation du verbe lui-même ; les représentations du verbe et du SP sont dans ce cas mutuellement dépendantes : le verbe apporte la cible et le SP le site. Le second mode de composition, illustré par (52), est l’ajout à la représentation du verbe ou du complexe verbal de conditions supplémentaires sur la localisation de la cible par rapport à un site introduit par le SP. La représentation du verbe est alors indépendante de celle du SP puisque la localisation a lieu par rapport à un site propre au SP ; le SP dépend du verbe pour la cible.

2.5.4 Conclusion

Ces travaux démontrent qu’une axiomatisation des verbes et prépositions spatiales est possible dans un domaine suffisamment étendu, ce qui pouvait sembler impossible *a priori* tant la richesse du langage naturel est grande.

On retiendra également de ces travaux le choix de la méthodologie employée qui se veut compositionnelle. Ainsi, Asher et Sablayrolles procèdent par construction incrémentale en partant du verbe puis en lui adjoignant les compléments de l’expression de changement de lieu. Les auteurs parviennent ainsi à donner une sémantique spatio-temporelle aux expressions de mouvement.

Voyons à présent une approche computationnelle dans le sens où elle se fonde sur un modèle mathématique muni de certaines heuristiques.

2.6 Le modèle de Hernández

Comme nous l’avons mentionné dans l’introduction, nos premiers efforts ont porté sur la réalisation d’une maquette informatique permettant de calculer les références spatiales portant sur un système d’affichage d’objets graphiques. Le modèle que nous allons présenter à présent qui est tiré de Hernández [1994] nous servira de base pour le module de calcul des référents spatiaux que nous développerons au chapitre 3.

Cette approche est d’inspiration fonctionnelle ; résolument qualitative, elle s’adapte finement au contexte comme nous allons le voir.

2.6.1 Les bases du modèle de Hernández

Après avoir insisté longuement sur les vertus d’une approche qualitative plutôt que quantitative, Hernández propose un modèle adapté à la prise en compte de la projection 2D d’une scène 3D (Hernández donne ensuite quelques prémisses et pistes pour une extension de ce modèle en 3D). Cette projection est une projection pure et simple, c’est-à-dire qu’on ne se soucie pas des éventuels problèmes de type “*devant/derrière*”, les objets ainsi projetés peuvent donc se chevaucher (voir la figure 2.10 par exemple).

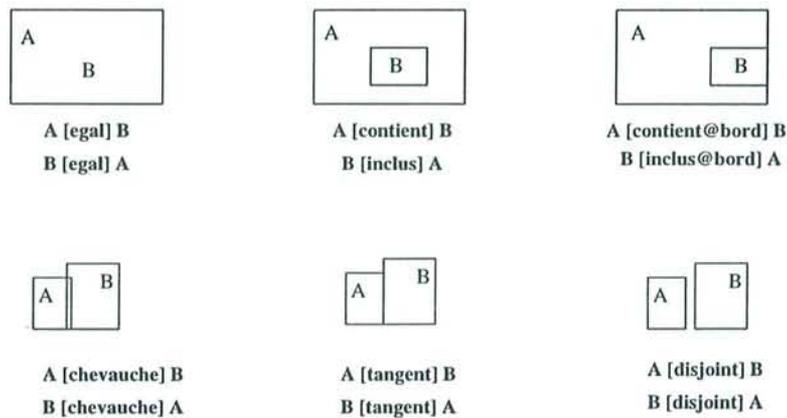


FIG. 2.10 – *Relations topologiques.*

Hernández travaille sur l’enveloppe convexe englobant les objets, ces derniers sont par ailleurs supposés être sans trous. D’après l’auteur, seuls deux facteurs suffisent à déterminer la position relative d’un objet par rapport à un autre : l’orientation des objets et leur extension. Il en résulte la définition de deux classes de relations spatiales disjonctives qui permettront de localiser précisément la position d’un objet par rapport à un autre à l’aide d’un doublet relation topologique/relation d’orientation :

- les relations topologiques (ignorent les problèmes d’orientation),
- les relations d’orientation (ignorent les problèmes topologiques).

2.6.2 Les relations topologiques

Hernández propose les huit relations topologiques binaires suivantes qui sont illustrées à la figure 2.10 :

- disjoint,
- tangent,
- chevauche,
- contient sur un bord,
- inclus sur un bord,
- contient,
- inclus,
- égal.

Ces différentes relations sont spécifiées dans le tableau ci-dessous où \bar{A} désigne l'extérieur de A au sens topologique du terme et \dot{A} désigne l'intérieur de A :

Relation	$\bar{A} \cap \bar{B}$	$A \cap \bar{B}$	$\bar{A} \cap B$	$A \cap B$
disjoint	\emptyset	\emptyset	\emptyset	\emptyset
tangent	$\neg\emptyset$	\emptyset	\emptyset	\emptyset
chevauche	$\neg\emptyset$	$\neg\emptyset$	$\neg\emptyset$	$\neg\emptyset$
contient sur un bord	$\neg\emptyset$	$\neg\emptyset$	$\neg\emptyset$	\emptyset
inclus sur un bord	\emptyset	$\neg\emptyset$	$\neg\emptyset$	\emptyset
contient	$\neg\emptyset$	$\neg\emptyset$	\emptyset	$\neg\emptyset$
inclus	\emptyset	$\neg\emptyset$	\emptyset	$\neg\emptyset$
égal	$\neg\emptyset$	$\neg\emptyset$	\emptyset	\emptyset

2.6.3 Les relations d'orientation

Au sein des relations d'orientation, Hernández considère tout d'abord des objets sans extension et propose de projeter sur l'objet de référence l'un des systèmes d'axes de la figure 2.11, ce système d'axes pouvant éventuellement subir une rotation afin de prendre en compte les orientations intrinsèque, déictique et extrinsèque¹⁹ comme nous le verrons dans un exemple par la suite.

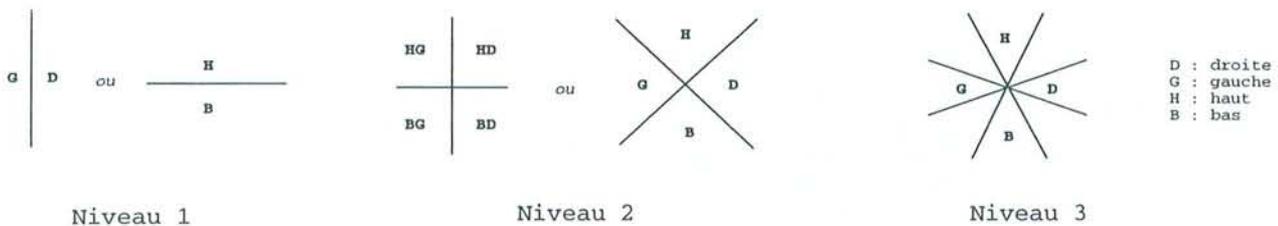


FIG. 2.11 – *Le modèle de Hernandez.*

Pour éclaircir le choix entre les différents systèmes d'axes, imaginons que l'on souhaite localiser deux cibles c_1 et c_2 par rapport à un site donné ; si on obtient la même préposition pour localiser les deux cibles au niveau i (deux choix de système d'axe sont parfois possibles, cf. le "ou" de la figure 2.11), il suffit de passer au niveau $i+1$ et ainsi de suite. Si arrivé au niveau 3, il y a encore ambiguïté, Hernández pense que plutôt qu'affiner encore le système d'axes, il convient de changer de site. Cette approche permet donc de prendre finement en compte le contexte en ne travaillant qu'au niveau de granularité nécessaire et suffisant.

Hernández adhère à la thèse selon laquelle seulement trois composantes sont nécessaires à l'établissement d'une relation d'orientation :

- l'objet de référence (appelé le site chez Vandeloise),

19. Ces trois orientations ont été définies ci-dessus pour le modèle de Vandeloise.

- l'objet référencié (appelé la cible chez Vandeloise),
- un cadre de référence qui peut être basé au choix sur l'orientation déictique, intrinsèque ou extrinsèque.

Finalement, la position relative de deux objets est décrite à l'aide de la clause : «*primary-object*²⁰[*topological, orientation*],*ref-object,ref-frame* » dans laquelle le *ref-frame* peut être de deux types :

1. implicite: c'est le cas lorsque cible et site tirent leur orientation de l'entité dans laquelle ils se trouvent inclus.
2. explicite: le cadre de référence est dans ce cas défini par la clause : «*type-of-use, [dex-relation [trans]]* » dans laquelle *type-of-use* doit être choisi parmi les trois orientations classiques : intrinsèque, déictique ou extrinsèque. Si *type-of-use* n'est pas du type intrinsèque, *dex-relation* contient alors la relation qui relie le site à un facteur externe ou à l'observateur. Enfin, une transformation supplémentaire peut être requise : *trans* pour prendre en compte les problèmes d'orientation en miroir et en tandem²¹.

2.6.4 Prise en compte des objets possédant une extension spatiale

Hernández démontre tout d'abord qu'on ne peut pas se contenter des centres de gravité pour décrire précisément la position d'un objet A par rapport à un objet B. Il suffit, pour s'en persuader, de prendre l'exemple de la figure 2.12 pour laquelle la seule prise en compte des centres de gravité aboutirait à situer B comme étant à droite de A dans les deux cas de figure, ce qui est faux dans le premier cas.

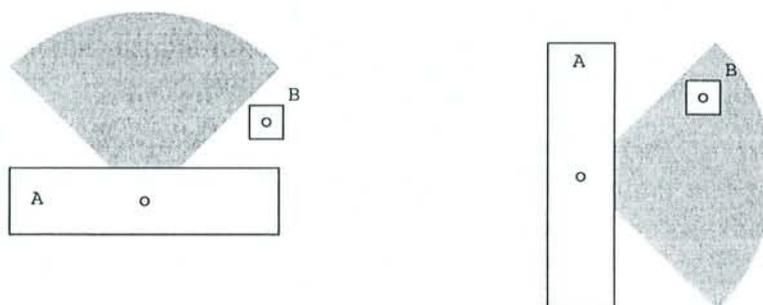


FIG. 2.12 – Les secteurs d'acceptabilité.

Hernández distingue donc trois cas selon que la cible et le site se trouvent *distants* l'un de l'autre, en situation de *proximité* ou finalement qu'ils se *chevauchent*. La cible est dite à proximité du site si son centre de gravité se situe en deçà de trois fois le diamètre du site. Conscient du caractère *ad hoc* de cette valeur, Hernández la justifie par le fait qu'elle fonctionnait bien pour l'application considérée que nous relaterons au paragraphe 2.6.5. Par ailleurs, Hernández précise que cette notion de proximité dépend également d'aspects

20. *Primary-object* se traduit par cible dans le formalisme de Vandeloise et *ref-object* par site.

21. Ces orientations ont été illustrées à la figure 2.4.

fonctionnels comme le temps ou l’effort qu’il faudrait fournir pour aller du site à la cible (il rejoint ici les aspects fonctionnels développés par Vandeloise [1986] pour les prépositions “près de/loin de”).

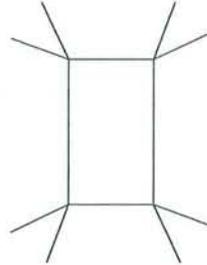


FIG. 2.13 – Zones d’acceptabilité non homogènes.

Hernández définit ensuite les algorithmes associés aux trois types de distances considérées :

- *pour les objets distants*: il suffit de tester dans quel secteur se trouve le centre de gravité de la cible, chaque secteur étant construit relativement au centre de gravité du site, les secteurs ne se chevauchant pas (figure 2.12 pour les secteurs associés à la préposition “au-dessus de” et “à droite de”).
- *pour les objets en proximité*: Hernández renvoie aux travaux de [Kobler, 1992]²² et précise simplement que là encore, les zones d’acceptabilité sont étendues pour les côtés et réduites pour les coins. Pour les cas ambigus (lorsque la cible se trouve à cheval sur deux zones d’acceptabilité) Hernández propose de recourir au système d’axes complémentaire du même niveau i , si l’ambiguïté demeure, il propose de revenir au niveau $i-1$ et de se poser le même problème.
- *pour les objets se chevauchant*: le facteur déterminant est la longueur du côté du site qui fait face à la cible. Il en découle que les coins doivent être traités séparément, ce qui conduit à l’obtention de zones d’acceptabilité non homogènes (figure 2.13). Hernández indique que ces zones sont obtenues en modifiant le modèle établi pour les objets distants. Sans donner plus de précision, l’auteur mentionne cependant que pour les objets étendus tels que les routes et les rivières on peut se contenter des prépositions de niveau 1 comme “à droite de/à gauche de” et “devant/derrière”.

2.6.5 Illustration

Voici un petit exemple pour illustrer le fonctionnement du modèle de Hernández fondé sur la dualité relation d’orientation/relation topologique (figure 2.14 où la partie pointillée désigne le devant intrinsèque de chaque bureau) :

22. [Kobler92] D. Kobler, *Visualisierung qualitativer Repräsentationen räumlichen Wissens*. Diplomarbeit, Institut für Informatik, Technische Universität München.

Le bureau de Stéphanie (A) se trouve à gauche du bureau de Daniel (B). (Figure 2.14 sans point de vue explicite). (53)

Vu de la bibliothèque qu'occupe Christian (D), le bureau de Stéphanie (A) se situe derrière celui de Daniel (B). (Figure 2.14 vue I). (54)

D'après ce qui précède, ces deux phrases sont traduites dans le formalisme de Hernández respectivement par (55) et (56) :

$$\text{REL1} = \langle A, [\text{disjoint, gauche}], B \rangle \quad (55)$$

$$\begin{aligned} \text{REL1} &= \langle A, [\text{disjoint, derrière}], B, \{\text{déictique, REL2}\} \rangle \\ \text{REL2} &= \langle D, [\text{disjoint, derrière}], B \rangle \end{aligned} \quad (56)$$

Étant donné que dans la phrase (53) aucun point de vue n'est donné *a priori*, c'est l'orientation implicite du bureau tout entier qui prévaut. Cette orientation induit que le bureau A se situe à gauche du bureau B. Pour la phrase (54), un point de vue explicite est mentionné. Selon Hernández, il faut donc introduire un *ref-frame* explicite qui va préciser la position du point de vue, et ce, relativement à l'objet de référence, (D se trouve derrière B).

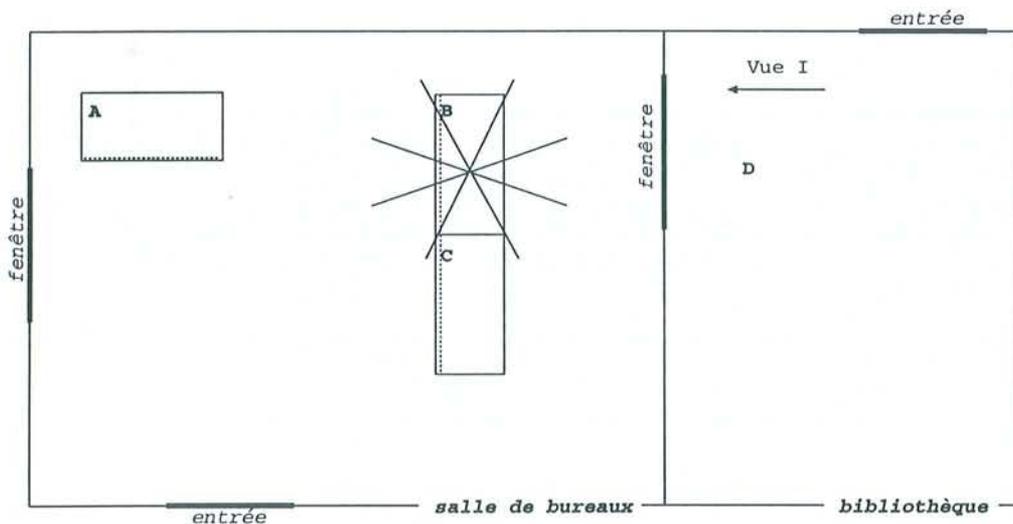


FIG. 2.14 – Prise en compte des différentes orientations dans le modèle de Hernández.

2.6.6 Conclusion

La contribution de Hernández présente une certaine originalité car l'auteur propose de prendre en compte le contexte en ayant recours à trois niveaux de granularité. Nous nous inspirerons de cette particularité au chapitre 3 afin d'élaborer notre propre modèle.

Nous tenterons cependant d'unifier ces trois niveaux ainsi que leurs éventuels sous-niveaux distingués par Hernández.

Le point délicat concernant les algorithmes de Hernández provient sans doute du caractère *ad hoc* de certaines valeurs utilisées (revoir en particulier la distinction faite entre les objets placés en situation distante, en situation de proximité et enfin en situation de chevauchement) qui nuisent à la méthodologie proposée. De notre côté, nous serons également confrontés aux aspects *ad hoc* de certaines valeurs que nous proposerons lors de l'élaboration des algorithmes qui seront présentés au chapitre 3.

Ces valeurs *ad hoc* ainsi que la caractère sans doute trop restreint des domaines d'application du modèle de Hernández nous font entrer dans la problématique plus générale de l'opposition entre les modèles généraux et les modèles spécifiques. Tandis que les premiers manquent de précision dans leur comportement, ce qui nécessite une méthodologie d'instanciation, les seconds sont bien souvent difficiles à étendre.

Nous retiendrons également de ces travaux, que, de par leur inspiration qualitative, ils rejoignent le principe d'économie cognitive que nous défendons au paragraphe 1.4.1.

2.7 Synthèse

Nous l'avons vu, il n'y a pas de relation biunivoque entre une représentation géométrique ou logique et une relation linguistique provenant de l'extrait d'un dialogue. Bien au contraire, le locuteur, son interlocuteur, mais également la tâche ont un rôle décisif à jouer.

Après avoir esquissé les limites des modèles d'inspiration purement géométrique, nous avons pu insister sur l'importance de la prise en compte des aspects fonctionnels. Ces derniers devraient tenir à l'avenir un rôle de plus en plus prépondérant au sein des approches linguistiques, logico-computationnelles et computationnelles.

Outre la notion de fonctionnalité qui joue le rôle de fil conducteur dans les différents modèles exposés, trois thèmes émergent également de ce chapitre. Il s'agit premièrement du choix en faveur des modèles d'inspiration qualitative par opposition aux modèles quantitatifs. Hernández penche très nettement en faveur des premiers comme cela apparaît dans ses travaux. Dans un second temps, nous avons souligné l'importance de la prise en compte non seulement du locuteur mais également de son interlocuteur. Enfin, nous avons également démontré que la tâche dans laquelle se situaient les deux locuteurs avait également un rôle essentiel à jouer (revoir en particulier les prépositions "*près de*" et "*loin de*" pour l'exemple de l'aviateur).

Chapitre 3

Le problème de la référence

Sommaire

3.1	Introduction	72
3.2	Le problème de la référence en linguistique	73
3.3	Le traitement de la référence en informatique	74
3.3.1	Prolégomènes	74
3.3.2	Le problème de la référence aux actions	76
3.4	Le problème de la référence aux objets en informatique	78
3.4.1	La DRT de Kamp	78
3.4.2	Le focus de Sidner	81
3.4.3	Un modèle de calcul des référents fondé sur la notion d'axiologie	82
3.5	Problématique et intérêt de la référence spatiale	88
3.5.1	Pourquoi s'intéresser aux références spatiales purement langagières?	89
3.5.2	Notre contribution	91
3.5.3	Conclusion	97
3.6	Une étude détaillée de la préposition "à gauche de"	98
3.6.1	Un modèle discrétisant l'espace	98
3.6.2	Un modèle linguistique	99
3.6.3	Un modèle à base de nuages d'acceptabilité	100
3.6.4	Un modèle fondé sur des inclusions de portions de surface	102
3.6.5	Un modèle à base de zones d'acceptabilité	103
3.6.6	Le modèle de Hernández	103
3.6.7	Autres limites de ces modèles	103
3.6.8	Conclusion	105
3.7	Vers une autre approche	105

3.7.1	Introduction	105
3.7.2	Rôle possible de la notion de saillance	106
3.7.3	Rôle possible de la notion de prototype	106
3.7.4	Vers un modèle de calcul des référents spatiaux	109
3.8	De la nécessité d'introduire la notion de cadre	121
3.9	Conclusion	122

3.1 Introduction

L'une des facilités essentielles du langage est de permettre la désignation variée d'objets et entités du monde. Ceci donne cohésion et continuité aux textes et discours, en procurant des mécanismes permettant d'éviter de répéter des informations déjà introduites [Sabah, 1989]. Un problème non trivial consiste alors à retrouver la ou les entité(s) que désignent ces reprises (auxquelles elles se réfèrent), d'où la notion de référence que [Kleiber, 1981] définit comme «*la fonction qui permet aux signes linguistiques de renvoyer à la réalité extra-linguistique*». Il s'agit d'un vaste problème qui a mobilisé bien des esprits en linguistique [Milner, 1982], [Kleiber, 1981], [Corblin, 1987] et [Corblin, 1995] pour ne citer qu'eux. En ce qui concerne l'informatique, comme l'a montré [Gaiffe, 1992], le problème de la référence limite beaucoup les performances des systèmes de dialogue actuels. En effet, on ne dispose pas jusqu'à présent de système unifié qui modéliserait le problème de la référence globalement, au contraire, on assiste à un foisonnement d'études qui s'intéressent à tel ou tel aspect particulier de la référence. Cette pluralité des modèles complique singulièrement la tâche de l'informaticien.

Après une introduction générale à la problématique de la référence en linguistique, nous verrons comment ce problème a été abordé en informatique, en insistant sur le modèle qui a été développé par Gaiffe et Romary [1993].

La seconde partie de ce chapitre se tournera vers le problème de la référence spatiale en l'illustrant sous l'angle spécifique du pilotage d'un système affichant des objets graphiques. Le fait que nous nous soyons restreint dans un premier temps à ce type d'application provient de l'étude bibliographique que nous avons réalisée. Cette dernière a démontré que les modèles développés jusqu'à présent, comme ceux de [Mukerjee, 1989], [Wazinski, 1991], [Briffault, 1992], [Schirra, 1990] et [Hernández, 1994], ne s'appliquaient qu'à des domaines très restreints (en particulier, leurs approches ne gèrent pas le cas des références plurielles comme nous le verrons dans ce chapitre). Au vu de ce constat, notre préoccupation a été dans un premier temps d'élaborer un modèle qui puisse remédier aux différents inconvénients des approches existantes dans le but d'unifier ces dernières pour parvenir, dans la suite de cette thèse, à généraliser ce modèle à des domaines plus étendus.

3.2 Le problème de la référence en linguistique

Tout au long de ce chapitre, nous nous appuyerons sur les notions d'anaphore, de déixis et de co-référence que l'on définit comme suit :

Anaphore : on parle d'*anaphore* lorsqu'un terme, souvent un pronom de troisième personne, est utilisé pour reprendre une autre expression nominale qui le précède, appelée traditionnellement son antécédent, à laquelle il emprunte sa référence, c'est-à-dire l'objet qu'elle désigne. Dans l'exemple de "*Pierre est ivre. Il a bu du cognac*", "*Pierre*" est l'antécédent du pronom "*il*", "*il*" est anaphorique et l'un et l'autre désignent le même objet, à savoir l'individu dans le monde qui souffre d'ébriété. Lorsque l'antécédent suit l'anaphore, on parlera de *cataphore* : "*Il a bu du cognac et il est ivre Pierre*".

Déixis : on parle de *déixis* ou d'*indexicalité* pour désigner les pronoms de première et de seconde personne, certains adverbes de temps comme "*maintenant, aujourd'hui, hier, demain, etc.*", et certains adverbes de lieu comme "*ici*". Ce que tous ces termes, appelés déictiques, ont en commun, c'est qu'on peut leur attribuer une signification, sur la base des indications linguistiques qui leur sont attachées, si l'on connaît la situation d'énonciation : "*je*" désigne la personne qui parle, "*tu*", la personne à qui l'on parle, "*maintenant*", le moment où l'on parle, etc. Lorsque l'on ignore la situation d'énonciation, il devient impossible d'attribuer toute la signification aux déictiques, comme le montre l'exemple suivant, emprunté à François Weyergans : "*Je suis ici aujourd'hui. Où serai-je demain pour dire que c'était hier?*".

Co-référence : deux termes sont co-référentiels s'ils désignent le même objet dans le monde. Par exemple, "*Jacques Chirac*" et "*le président de la République française en 1996*" sont co-référentiels.

La notion de monde de référence ou de contexte a sans aucun doute un rôle décisif à jouer dans le calcul des référents. Ainsi, si je dis "*le président*" sans avoir posé auparavant un contexte de référence, cette référence sera nécessairement ambiguë. Le monde de référence est donc déterminant dans la contextualisation du calcul référentiel. Ce point fait précisément partie de l'une des contributions de cette thèse où nous cherchons à cerner plus finement le contexte pour le cas des énoncés spatiaux et ce à l'aide de la notion de cadre de référence.

Par ailleurs, en linguistique, on distingue généralement cinq types de références : la référence *directe* (57), la référence *indirecte* (58), la référence *démonstrative* (59), la référence *déictique* (60) et la référence *anaphorique* (61) dont voici respectivement un exemple de chaque type tirés de [Moeschler et Reboul, 1994] :

Marco Polo est aussi un personnage des Villes invisibles (57)
d'Italo Calvino.

L'omelette au jambon est partie sans payer. (58)

Cet enfant a de la fièvre. (+ geste). (59)

J'ai mal aux dents. (60)

Pierre a perdu son chapeau. Il est très distrait. (61)

En (57), les noms propres “*Marco Polo*” et “*Italo Calvino*” servent à effectuer des actes de référence directe ; en (58) la description définie (article défini + nom (+ modificateur)) “*l'omelette au jambon*” sert à effectuer un acte de référence indirecte ; en (59) le groupe nominal démonstratif “*cet enfant*” sert à opérer un acte de référence démonstrative ; en (60) le pronom de première personne “*je*” sert à effectuer un acte de référence déictique ; en (61) le pronom de troisième personne “*il*” sert à effectuer un acte de référence anaphorique.

Les cinq types de référence que nous venons de mentionner relèvent en fait plus de différents points de vue sur le problème de la référence que d'un système parfaitement établi. En effet, il n'y a pas toujours de relation biunivoque entre la forme des énoncés et le type de référence considéré. À titre d'illustration, il suffit de prendre les phrases (62) et (63), où le pronom “*elle*” joue le rôle de référent anaphorique en (62) et celui de référent déictique en (63) si l'on imagine par exemple pour la phrase (63)¹ que vous vous trouviez dans un aéroport avec un ami et que vous regardiez une personne qui essaie de rentrer une valise de taille démesurée dans sa voiture.

*Mireille a de la fièvre, elle ne se sent pas très bien
aujourd'hui.* (62)

Regarde, elle ne rentrera jamais. (63)

Le fait qu'il n'y ait pas toujours bijection entre la forme et l'un des cinq types de cas référentiel cités ci-dessus doit être nuancé pour le cas de la référence spatiale. Il semble en effet, quelle que soit la forme prise par l'énoncé, que ce dernier fait toujours intervenir une composante déictique. De notre point de vue, cela provient sans doute de la nature des référents spatiaux qui doivent nécessairement se fonder sur une portion de l'espace pour prendre leur signification.

Il nous semble à présent nécessaire d'étudier la manière dont les problèmes référentiels ont été abordés en informatique.

3.3 Le traitement de la référence en informatique

3.3.1 Prolégomènes

Dans le domaine du dialogue homme-machine, et plus particulièrement quand il s'agit de dialogues multimodaux, le problème de la résolution des références peut être abordé de deux manières différentes. Dans le cas où les utilisateurs du système sont

1. Exemple tiré de [Kleiber, 1990b].

des professionnels prêts à accepter une période de formation ou lorsque l'application est suffisamment simple, l'approche la plus efficace est de définir un langage artificiel tel que celui qui a été mis en œuvre par Souvay [1992] pour la commande d'une console sonar à partir du système DIAPASON. Dans de telles situations, le langage est défini de manière à simplifier la résolution des références exprimées par l'utilisateur. Ainsi, dans le système DIAPASON, le système informatique s'appuie sur la forme des énoncés qui est du type *prédicat + argument(s)* pour résoudre les anaphores et les ellipses. Voici un exemple de dialogue tiré de [Souvay, 1992] pour le cas d'un système informatique dédié à la commande d'une console sonar :

Utilisateur : *Libérer mémoire VPR.*
 Ordinateur : *Confirmez : libérer mémoire VPR0.*
 Utilisateur : *Négatif VPR.*
 Ordinateur : *Confirmez : libérer mémoire VPR.*
 Utilisateur : *Ok.*
 Ordinateur : *Ok.*

Si les utilisateurs ne sont pas des professionnels, il n'est pas possible de leur imposer d'utiliser des expressions ayant une structure syntaxique rigide ou simplifiée sans les contraindre fortement. Les études de Mignot [1995], Pierrel et Romary [1997] ont montré que même si l'utilisateur se situe dans une tâche très finalisée, le langage qu'il utilise correspond à un sous-langage² de ce qu'on nomme communément le langage naturel. Ainsi, il s'avère que quelle que soit la tâche considérée, il n'est pas possible d'obtenir de la part d'un utilisateur une réduction de la palette des expressions référentielles dont il dispose de façon générale, à savoir les expressions définies, indéfinies, démonstratives ou pronominales. Les restrictions observées se situent beaucoup plus au niveau du nombre de concepts pertinents pour la tâche et se traduisent par une réduction du lexique et des valeurs sémantiques que peuvent prendre les entrées dans celui-ci. L'option que nous avons choisie est donc de considérer les expressions référentielles telles qu'elles sont utilisées dans le langage courant et de parier qu'elles seront employées d'une façon similaire en situation de dialogue finalisé, plutôt que de préjuger des réactions d'un utilisateur devant un système de dialogue et limiter nos implémentations à de simples projections (qui bien souvent s'apparentent plus à des absurdités³).

Pierrel et Romary [1997] distinguent trois grandes classes d'applications qui, en fonction du type d'interaction entre l'utilisateur et le système, mettent en œuvre différents

2. Pierre Falzon [1986] parlerait de *langage opératif*. Les langages opératifs sont plus faciles à modéliser d'un point de vue informatique. Par ailleurs, les mécanismes qui les sous-tendent n'entrent jamais en contradiction avec ceux qui sont mis en œuvre au sein de la langue naturelle.

3. C'est notamment le cas des démonstratifs qui étaient, encore très récemment, presque systématiquement traités de manière co-référentielle alors que la reprise anaphorique existe :

- 1) *Déplace la fenêtre verte.*
- 2) *Mets cette fenêtre en arrière plan.*

processus de co-référenciation :

- **le transfert ou la saisie d'information** : le système de dialogue se limite alors à un rôle d'intermédiaire permettant de remplacer par la langue l'usage d'un langage formel de communication avec une base de données ou de connaissances. Le plus souvent, cette saisie d'informations se réalise sans véritable interprétation des données langagières et le traitement des références n'y est le plus souvent pas abordé sauf sous les aspects de traitement des quantificateurs dans des systèmes où le langage cible est du type SQL.
- **la commande (ou coordination) d'action** : le système de dialogue doit alors jouer un rôle d'intermédiaire ou d'assistance auprès de l'utilisateur dont le but est la commande d'un système automatisé ou informatique. Il présuppose donc, du moins dans les systèmes actuellement envisagés, un univers applicatif entièrement modélisable et l'un des objectifs principaux du dialogue utilisateur-système est de déterminer *que faire* (références aux actions) et *sur quoi* (référence aux objets).
- **l'acquisition de connaissances** : que ce soit par le système (apprentissage automatique) ou l'utilisateur (enseignement ou formation assistée par ordinateur), il ne s'agit plus alors uniquement de sélectionner des co-références communes, mais de construire petit à petit chez l'interlocuteur un modèle de référence commun avec son interlocuteur. Un rôle particulier dans ce cas est dévolu aux mécanismes d'argumentation indispensables à toute tâche de ce type.

De notre côté, nous nous intéresserons à la seconde grande classe d'applications c'est-à-dire aux dialogues de commande pour lesquels nous tenterons de démontrer qu'ils nécessitent des traitements fins de la référence qui ont été, par trop souvent, sous-estimés.

3.3.2 Le problème de la référence aux actions

On distingue traditionnellement la référence aux actions de la référence aux objets. Tandis que la première vise à déterminer "*que faire ?*", la seconde s'intéresse au problème de savoir "*sur quoi ?*".

Duermael [1994] démontre dans sa thèse, que la plupart des systèmes actuels opèrent une bijection trop rigide entre les énoncés émis par l'utilisateur et les opérateurs élémentaires ou complexes dont dispose le gestionnaire ou le modèle de l'application informatique. Il en ressort que les structures langagières acceptées par l'ordinateur se retrouvent directement liées aux opérateurs informatiques dont dispose ce dernier. Par ailleurs, la bijection entre énoncés et opérateurs entraîne la définition d'un langage artificiel dont la complétude et la régularité langagière n'est pas acquise. Duermael souligne également le fait que l'ordre des opérations informatiques à effectuer ne respecte pas forcément l'ordre des énoncés [Duermael, 1994] p. 79.

Ces réflexions conduisent l'auteur à proposer une modélisation de la référence aux actions en terme d'état initial et d'état final, c'est-à-dire qu'un énoncé de commande spécifie la situation finale à atteindre compte tenu de la situation courante, la sémantique du prédicat se réduit alors à la modification d'une ou plusieurs propriétés des objets en vue d'atteindre un état prédéterminé. Ce point de vue nous rapproche du schéma classique [pré-conditions] action(s) [post-conditions] qui est utilisé en vérification des programmes.

On peut finalement synthétiser la démarche de Duermael selon trois étapes :

Une analyse syntaxico-sémantique qui aboutit à la construction d'une structure de l'énoncé en terme de syntagmes verbaux, groupes nominaux et groupes prépositionnels.

Une analyse linguistique plus fine, ou compréhension, qui fournit en fonction de l'état de la tâche et de la structure de l'énoncé une spécification d'un état initial et d'un état final.

La phase de planification qui permet d'engendrer les commandes élémentaires à exécuter. Duermael a recours en fait à un planificateur qui fonctionne en chaînage arrière. Le principe sous-jacent à ce type de chaînage correspond à rechercher, à partir de l'état final, quels sont les objets et propriétés qui diffèrent de l'état courant ou initial, puis de déduire par filtrage sur les post-conditions des opérateurs de l'application, quels sont ceux qu'il faut planifier. Pour un exposé plus détaillé concernant le chaînage arrière, voir [Haton et Haton, 1990].

Le principal intérêt des travaux de Duermael réside dans la transformation du problème de la référence aux actions en un problème de référence aux objets. En effet, l'énoncé est vu comme un but à atteindre et se traduit alors comme une modification particulière à réaliser sur l'ensemble des objets de la tâche. Ainsi, prenons l'exemple de la phrase "*place le triangle à droite du cercle*" qui serait énoncée à la figure 3.1.(a). Duermael transformerait ce problème en celui de la figure 3.1.(b) où le triangle occuperait la nouvelle position "*à droite du cercle*", position qu'il faudrait bien entendu calculer à l'aide d'une référence spatiale.

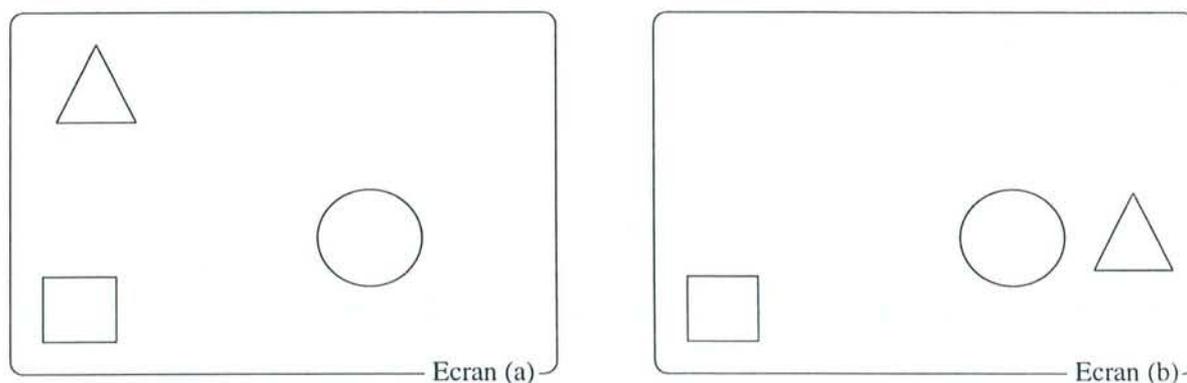


FIG. 3.1 – *Le problème de la référence aux actions.*

Avant d'étudier le modèle qui a été développé par Gaiffe et Romary, intéressons-nous à quelques modèles qui relèvent prioritairement du traitement de la référence aux objets.

3.4 Le problème de la référence aux objets en informatique

3.4.1 La DRT de Kamp

a) Présentation du modèle

La DRT (Discourse Representation Theory) de Kamp [1988], [1993] se place en continuité des travaux de Montague [1970]⁴. Tandis que l'approche montagovienne ne modélise que les aspects référentiels intra-phrastiques, les travaux de Kamp visent à saisir la cohésion discursive en assurant le suivi des relations d'équivalence. Dans le cas de la DRT, les structures sont des DRSs (Discourse Representation Structures). Les DRSs sont construites à chaque énoncé sur la base de critères syntaxiques, c'est-à-dire des informations contenues strictement dans l'énoncé. Elles contiennent les référents introduits par le discours et les prédicats qui s'y appliquent.

Ainsi, l'énoncé :

Jean offre des fleurs à Marie. (64)

conduit à la DRS suivante :

x, y, z
Jean (x)
Marie (y)
fleurs (z)
offrir-à (x, y, z)

où la partie supérieure contient les éléments constituant l'univers du discours (en tant qu'entité construite ou reprise par le discours) et la partie inférieure les conditions de validité de la DRS construite, sous la forme de prédicats vérifiés par ces éléments.

Les énoncés suivants, s'ils poursuivent le discours, vont contribuer à la construction de nouvelles DRSs qui viennent s'ajouter aux précédentes. Dans le cas contraire, on a une rupture dans le discours, qui conduit à une nouvelle succession de DRSs. Par exemple, si (64) est suivi de :

Elle les trouve belles. (65)

on construira dans un premier temps la DRS correspondante, qui représente le contexte

4. On pourra se reporter à [Dowty *et al.*, 1981] pour une bonne introduction aux grammaires de Montague.

du discours local à l'énoncé :

t, v
t=?
v=?
trouver-belles (t, v)

avant de rattacher les référents à ceux de la DRS précédente, en indiquant une égalité entre les variables. Dans la représentation de l'énoncé (65) les égalités incomplètes indiquent que les référents préexistent à l'énoncé, et qu'ils doivent donc être retrouvés dans la DRS précédente puisqu'on n'envisage ici que les liens anaphoriques. Les contraintes sur le genre et le nombre entre les référents de (64) et les pronoms de (65) permettent alors de poser les égalités, en conduisant à la nouvelle DRS qui sera considérée pour la suite du discours.

x, y, z, t, v
Jean (x)
Marie (y)
fleurs (z)
offrir-à (x, y, z)
t=y
v=z
trouver-belles (t, v)

Par la construction des DRSs, la DRT permet une résolution co-référentielle dès lors que les référents sont introduits par le discours. Elle constitue aussi de fait (d'où son intitulé) une représentation de ce dernier.

L'un des mécanismes essentiels sur lequel se fonde la DRT concerne la modélisation du domaine de validité des référents. En particulier, il est possible de former des espaces inclus dans l'espace de départ tels que les variables qui y sont incluses deviennent inaccessibles. En s'appuyant toujours sur la syntaxe, la DRT propose des règles pour traiter la négation, l'implication, la disjonction... Ainsi, à titre d'exemple, considérons la négation suivante :

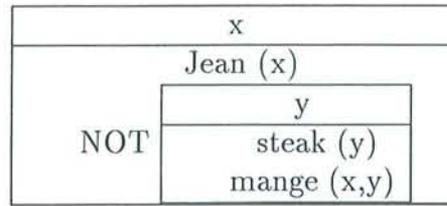
Jean ne mange pas de steak. (66)

Il mange une choucroute. (67)

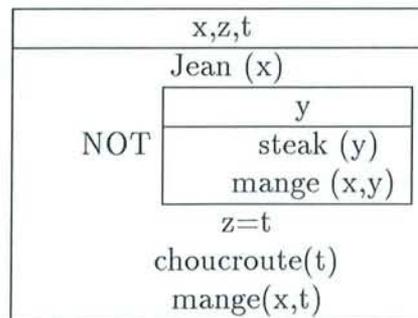
Ici, quel que soit ce qui suit le pronom "il", "il" ne peut pas reprendre le "steak" et ce indépendamment de toute considération sémantique. La DRS correspondant au premier énoncé sera dans un premier temps :

x
Jean (x)
pas-manger-un-steak (x)

La négation fait construire une DRS emboîtée dans la DRS de départ, on obtient donc :



L'énoncé suivant s'insère dans la DRS la plus englobante. On introduit une variable z pour le pronom, mais on ne peut pas avoir $z = y$ car y n'est pas accessible dans la DRS englobante. On obtient finalement :



Par des règles du même type, la DRT traite le problème de la quantification, celui de la sommation d'éléments, etc. Ces règles permettent donc, sur la base de la syntaxe, de créer des contextes rendant possible ou non la co-référence entre un pronom et un éventuel antécédent.

b) Bilan de la DRT

Un bref bilan de la DRT nous indique qu'au moins un point essentiel est problématique dans le cadre même du discours, et plus particulièrement du dialogue : l'univers du discours est clos par rapport à celui-ci. En effet, seuls deux types de références sont envisageables. D'une part, les co-références (le référent appartient alors nécessairement à l'univers du discours existant avant l'énoncé), d'autre part, l'indication d'un nouveau référent dans le discours, ce qui est le cas des variables x , y et z dans l'énoncé (66). Cela signifie que le contexte se réduit à l'univers du discours, ce qui est gênant, et tout particulièrement dans le cas d'un dialogue. En effet, la simple présence d'éléments visuels constitue également un contexte, et face à une interface graphique, par exemple, on peut tout à fait référer à des éléments représentés sur l'écran sans que ceux-ci aient été précédemment introduits dans (ou par) le discours.

D'autre part, étant donné que les DRSs sont construites à chaque énoncé sur la base de critères syntaxiques, l'approche de Kamp est essentiellement codique⁵ ce qui laisse peu de place (jusqu'à maintenant) pour la prise en compte de phénomènes plus pragmatiques (prendre par exemple le cas des implicatures^(*)).

5. Par "codique", nous entendons le fait que cette approche est très mécanique.

3.4.2 Le focus de Sidner

Tout comme les travaux de Kamp portant sur la DRT, l'approche de Sidner [1986] s'intéresse au problème des références intra-discursives. En particulier, partant du fait que le thème du discours, au sens habituel des entités dont parle le dialogue, a un rôle à jouer dans la résolution des anaphores, Sidner propose d'utiliser la notion de *focus* pour résoudre les anaphores définies, c'est-à-dire aussi bien les pronoms que les groupes nominaux déterminés par l'article défini. Comme le thème ne change pas de façon aléatoire, Sidner considère que le discours suit une forme arborescente. Autrement dit, les thèmes successifs s'organisent en thèmes et sous-thèmes locaux à ces thèmes. On pourra se reporter à [Sidner, 1986] ou [Sabah, 1989] p. 233-236 pour un exposé détaillé de l'algorithme qui gère les focus en ayant recours à une pile informatique.

L'une des critiques que nous formulerons à l'encontre de cette approche porte sur la gestion des focus. En effet, Sidner traite de la même manière l'anaphore par un pronom et l'anaphore par un groupe nominal, or on peut très facilement construire un énoncé qui démontre que l'anaphore par un groupe nominal peut fonctionner à une bien plus grande distance que l'anaphore par un pronom. Il semble alors désastreux de faire sortir à un instant donné de la pile un groupe nominal qui pourrait être repris dans la suite du discours. Pour illustrer ce propos, nous prendrons l'exemple suivant qui a été tiré de [Gaiffe, 1992] où il est très gênant de considérer que le groupe nominal "*la voiture*" fasse oublier (en le faisant sortir de la pile) le focus précédent "*le garagiste*" :

*J'ai acheté hier une voiture au garagiste de la rue en dessous de chez moi.
Il m'a fait vingt pourcents.
Je suis rentré chez moi.
La voiture ne roulait pas bien.
Je suis donc retourné le voir.
Il me l'a échangée contre une autre en meilleur état.*

Enfin, le principal reproche que nous ferons à cette approche provient du fait que Sidner ne s'intéresse qu'aux références intra-discursives. En effet tout comme la DRT, et ce, dans la perspective de l'accès à un univers d'objets visualisés à l'écran, on voit mal comment le modèle de Sidner pourrait croiser les référents provenant du discours et ceux provenant de l'espace visuel que fournit l'ordinateur. On pourra trouver dans [Gaiffe *et al.*, 1994] d'autres critiques du modèle de Sidner.

Les insuffisances du modèle de la DRT de Kamp et de celui du focus de Sidner ont conduit Romary et Gaiffe à mener une réflexion approfondie sur le problème de la référence que nous allons décrire à présent.

3.4.3 Un modèle de calcul des référents fondé sur la notion d'axiologie

a) Introduction

L'approche que nous allons présenter est le fruit d'une réflexion menée depuis quelques années par Romary et Gaiffe sur les problèmes de référence. Son aboutissement peut être retracé dans [Romary, 1989], [Gaiffe *et al.*, 1991], [Gaiffe, 1992], [Gaiffe *et al.*, 1992] et [Gaiffe et Romary, 1993], l'exposé le plus détaillé de ce modèle est effectué dans [Gaiffe, 1992].

Ces travaux se fondent sur la notion d'axiologie qui est un point de vue référentiel adopté par un discours pour exprimer une opposition élémentaire, par exemple: "blanc"/"noir"⁶. Cette opposition élémentaire doit être vue dans une perspective contrastive, c'est-à-dire qu'elle va permettre de faire reposer l'interprétation des expressions référentielles sur un mécanisme qui isole un référent, bien plus qu'il ne l'identifie de par ses caractéristiques propres [Romary et Gaiffe, 1995]. À titre d'illustration, il suffit de considérer par exemple une scène sur laquelle se trouvent affichés des cercles et un rectangle hachuré. Si l'utilisateur souhaite détruire le rectangle, il pourra très bien dire "détruis le rectangle" (l'utilisateur exploite au mieux les différences de type) plutôt que "détruis le rectangle hachuré" (l'utilisateur donne alors toutes les caractéristiques de l'objet même si ce n'est pas nécessaire). Mounin [1972] a également souligné l'importance de cette structuration binaire dont il pense qu'une analyse psychologique et linguistique plus fine pourrait démontrer qu'elle correspond peut-être à un caractère fondamental du fonctionnement de la pensée humaine.

Gaiffe et Romary se placent dans un premier temps dans la perspective de calculer le ou les référents d'un groupe nominal du type $\langle \text{Dét} \rangle \langle \text{Adjectif}_{\text{éventuel}} \rangle \langle \text{Nom} \rangle \langle \text{Adjectif}_{\text{éventuel}} \rangle$ qui serait énoncé par un utilisateur donné, placé devant un écran d'ordinateur, sur lequel se trouvent affichés un certain nombre d'objets. Pour ce faire, ils effectuent la distinction entre deux types de propriétés: les propriétés nominales (notées N) et les propriétés particulières (notées P). Les propriétés nominales regroupent tout ce qui, du point de vue du sens, est directement exprimé dans le nom commun et l'(les) adjectif(s) en tant que sous-catégorisant(s). Quant aux propriétés particulières, notées P, ce sont elles qui spécialisent un référent en dessous de propriétés N. Ce sont ces propriétés qui permettent d'isoler un élément de la classe de tous ceux qui sont N⁷. Pour résumer les caractéristiques des propriétés P, on remarquera qu'elles réalisent en fait une opération de catégorisation ou de sous-catégorisation tandis que les propriétés N permettent de passer de la catégorie ou de la sous-catégorie à l'instance. Voyons à présent les schémas référentiels associés aux différents types de groupes nominaux non génériques⁸.

6. La notion d'axiologie peut être rapprochée de la sémantique différentielle chère à [Rastier, 1987; Rastier, 1991].

7. Corblin [1987] avait déjà défendu bien avant la thèse de la recherche d'un référent donné, par discrimination d'une ou plusieurs propriétés au sein d'un contexte donné.

8. "Un triangle a trois côtés" est un exemple de groupe nominal générique en "un N" tandis que "l'homme est mortel" est un exemple générique en "le N". Comme le mentionne Gaiffe [1992], les groupes nominaux génériques n'apparaissent que très peu dans les dialogues de commande finalisés. Il convient

b) Les groupes nominaux en “le N”

Les groupes nominaux de la forme “le N” sont destinés à sélectionner dans un ensemble d’objets celui qui porte le nom N par opposition aux objets du même ensemble qui ne méritent pas ce nom. Le fait qu’il est souhaitable qu’il y ait un N et des $\neg N$ pour les groupes nominaux définis permet d’expliquer des exemples tels que (68) et (69) :

Déplace les fenêtres vertes. (68)

Mets la fenêtre la plus à gauche en arrière plan. (69)

Ainsi, dans la phrase (69), les deux ensembles candidats pour justifier le GN défini sont respectivement : l’ensemble de tous les objets visibles sur l’écran, ce qui est l’interprétation non anaphorique, et l’ensemble des fenêtres vertes, ce qui correspond à l’interprétation anaphorique. La présence du N “fenêtre” dans le second énoncé serait inutile s’il s’agissait d’une anaphore : dans l’ensemble des fenêtres vertes, tous les éléments sont des fenêtres ce qui nous ferait plutôt dire : “mets la plus à gauche en arrière plan” pour le cas d’une anaphore. En revanche, cette catégorie “fenêtre” se justifie pleinement dans le cas de l’interprétation non anaphorique : parmi les objets présents à l’écran, certains sont des fenêtres, d’autres non.

Dans le cas de l’exemple (70) et (71) l’ensemble correspondant à l’interprétation anaphorique devient un aussi bon candidat que l’ensemble correspondant à l’interprétation non anaphorique ; l’ambiguïté est donc possible. Il faudrait alors d’autres critères pour faire le bon choix.

Déplace les fenêtres vertes et les icônes bleues. (70)

Mets la fenêtre la plus à gauche en arrière plan. (71)

Finalement, les auteurs proposent le schéma de la figure 3.2 pour lequel on a bien un ensemble (noté Q) et une axiologie (N versus $\neg N$). Sur cette axiologie, on se situe au point noté : “N”. L’intérêt de ce schéma est qu’il permet de retarder le choix entre la référence directe et l’anaphore. Dans le premier cas, l’ensemble Q du schéma sera instancié par l’ensemble des objets de l’écran, dans le second, il s’agira d’un ensemble issu des énoncés précédents comme dans (72) et (73) où l’ensemble Q est instancié par {fenêtre-verte, icône-bleue}.

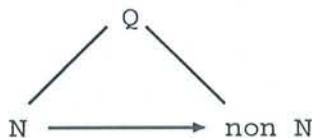


FIG. 3.2 – “le N”.

Déplace la fenêtre verte et l’icône bleue. (72)

malgré tout d’avoir ces exemples en tête afin que les modèles proposés pour les GN définis par exemple n’entrent pas en contradiction avec les cas où ceux-ci fonctionnent de manière générique.

Mets la fenêtre en arrière plan.

(73)

Un problème semble ne pas être pris en compte par ce modèle, ainsi si je dis “*détruis la fenêtre*”, il paraît naturel de rechercher une entité qui ait le type fenêtre par rapport à d’autres entités qui n’aient pas ce type (ce qui est bien pris en compte par le schéma de la figure 3.3.(a)).

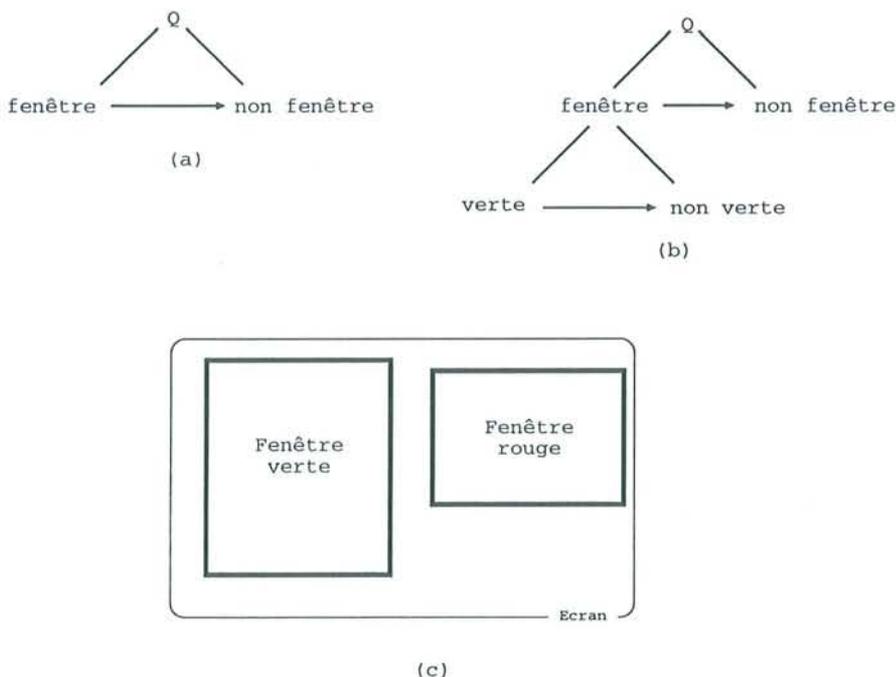


FIG. 3.3 – De l’intérêt de l’axiologie “fenêtre” versus \neg “fenêtre”.

Par contre, de “*détruis la fenêtre verte*” il semble curieux de conserver l’axiologie “fenêtre” \rightarrow \neg “fenêtre” (voir à la figure 3.3.(b)) en imaginant que seules une fenêtre verte et une fenêtre rouge se trouvent affichées sur l’écran (voir figure 3.3.(c)). En effet, il n’y a pas de \neg “fenêtres” affichées à l’écran ce qui rend impossible l’instanciation de l’axiologie “fenêtre” \rightarrow \neg “fenêtre”. Cependant, le fait de parler de \neg “fenêtres” est légitime car on pourrait aussi bien mentionner l’écran, le fond de l’écran, etc. Pour corroborer ce fait, imaginons un dialogue où l’utilisateur n’aurait que la possibilité de manipuler des fenêtres en ayant recours à des phrases du type: “*Détruis la fenêtre bleue. Mets une fenêtre verte ici, etc.*” Ainsi, s’il mentionnait dans chacune de ses requêtes le nom “fenêtre”, l’utilisateur apporterait trop d’information à la machine en ne respectant pas la maxime de Grice [1979] de quantité qui conduirait plutôt l’utilisateur à utiliser des requêtes du type: “*Détruis la bleue, mets une verte ici, etc.*”

c) Les groupes nominaux en “ce N”

“Ce N” permet de re-catégoriser un référent ([Gaiffe, 1992] p. 49-50). Le référent doit donc apparaître en dessous de la catégorie N. De ce fait, lorsqu’on dit “ce N”, on oppose un individu particulier, celui dont on parle aux autres qui sont également N. Cet individu

particulier, pour pouvoir être opposé aux autres N doit avoir des propriétés qui lui soient propres. On a donc le schéma de la figure 3.4. Les propriétés permettant d'individualiser le référent pourront être le fait d'être désigné du doigt ou d'avoir été singularisé dans le discours. Autrement dit, "ce N" ne pourra reprendre qu'un référent saillant.

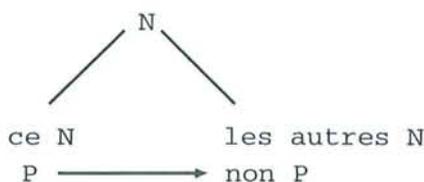


FIG. 3.4 – "ce N".

Pour "ce N" comme pour "le N", les schémas référentiels proposés unifient donc le cas de l'anaphore et celui de la référence directe.

d) Les groupes nominaux en "un N"

Dans le cas des groupes nominaux indéfinis, Romary et Gaiffe envisagent deux axes d'opposition: "un chat" fait partie de la classe des chats et peut donc s'opposer à ce qui n'est pas chat, c'est également un individu susceptible de se distinguer des autres chats. Ces deux axes d'opposition possibles nous amènent au schéma de la figure 3.5. La différence avec le groupe démonstratif tient au fait que l'indéfini permet d'introduire un nouveau référent dans le discours, l'axe P versus $\neg P$ qui individualise le référent dans la classe est donc posé *a priori* dans le cas de l'indéfini. Les auteurs justifient l'axe d'opposition N versus $\neg N$ à l'aide de l'exemple:

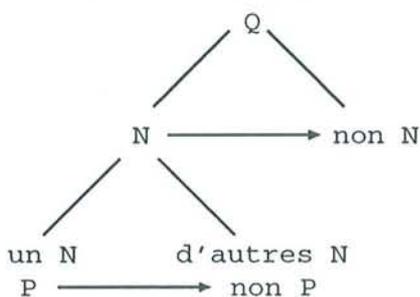


FIG. 3.5 – "un N".

J'ai acheté une Toyota parce qu'elles sont robustes et pas chères. (74)

Il semble clair qu'à partir de cet exemple, un fort sous-entendu porte sur le caractère moins robuste d'autres voitures que les Toyotas. C'est bien ce qui justifie de disposer à partir de l'antécédent "une Toyota" de la possibilité d'opposer la classe des Toyotas à d'autres classes. Cette opposition est précisément matérialisée par l'axe N versus $\neg N$ du schéma de référence de la figure 3.5.

Notons qu'on ne peut rassembler dans Q que des éléments qui sont relativement homogènes d'un point de vue sémantique car une axiologie n'a de sens que dans un certain ensemble qui permette de regrouper les objets que l'on oppose. Ainsi, si deux objets ne peuvent pas être rassemblés, les opposer n'a en définitive pas grand sens, il suffit de prendre pour exemple : "*Hier j'ai vu Paul et le marteau*"?! En résumé, si Q peut rassembler différentes catégories, ces dernières possèdent nécessairement une certaine cohérence sémantique.

e) Conclusion

Le principal avantage de ce modèle tient au fait que la classification qu'il propose n'est pas du type désignation directe, anaphore ou co-référence avec un geste mais qu'il opère plutôt selon la forme syntaxique des groupes référentiels considérés. Par ailleurs, les schémas référentiels de Gaiffe et Romary offrent une vision plus abstraite du monde représenté et traduisent le fait que la langue fournit bien souvent plus un point de vue sur des connaissances qu'une réelle description de celles-ci.

La principale limitation de ce modèle provient du fait qu'il paraît difficilement extensible, tel quel, au problème de la référence spatiale. En particulier, une référence comme celle qui apparaît à la phrase (75) nécessite, en regard de la phrase (76) de donner précisément une signification à l'axiologie "*à droite de*" \rightarrow \neg "*à droite de*" qui dépasse, en complexité, le fait d'être "*rouge*" ou \neg "*rouge*".

Détruis l'icône de droite. (75)

Détruis l'icône rouge. (76)

D'autre part, il devient très vite nécessaire d'ajouter à ce modèle des règles *ad hoc* afin que la notion d'axiologie puisse fonctionner correctement. En particulier, l'ordre d'application des axiologies n'est pas indifférent : il suffit de prendre le cas des références (77) et (78) où chacune de ces références est codée respectivement par les schémas 3.6.(a) et 3.6.(b). Si nous cherchons alors à appliquer chacun de ces deux schémas sur la figure 3.7, il apparaît immédiatement que le schéma 3.6.(b) aboutirait à un ensemble vide car l'axiologie "*de droite*" \rightarrow \neg "*de droite*" conduit au singleton "*icône verte*" que l'on ne pourra spécialiser davantage par l'axiologie "*rouge*" \rightarrow \neg "*rouge*".

Détruis l'icône rouge de droite. (77)

Détruis l'icône de droite, rouge. (78)

La différence entre les axiologies "*rouge*" \rightarrow \neg "*rouge*" et "*de droite*" \rightarrow \neg "*de droite*" provient du fait que la première met en jeu des propriétés possédant un caractère intrinsèque, c'est-à-dire qu'elles peuvent être directement évaluées sur le référent tandis que la seconde axiologie doit faire appel au contexte pour discriminer l'entité considérée⁹.

9. On peut certes imaginer des contextes où se trouveraient affichées un certain nombre d'icônes qui seraient plus ou moins rouges (avec des nuances par exemple). Une requête du type "*détruis l'icône rouge*" s'apparenterait alors à détruire l'icône qui serait la plus rouge. Ceci nous rapprocherait alors

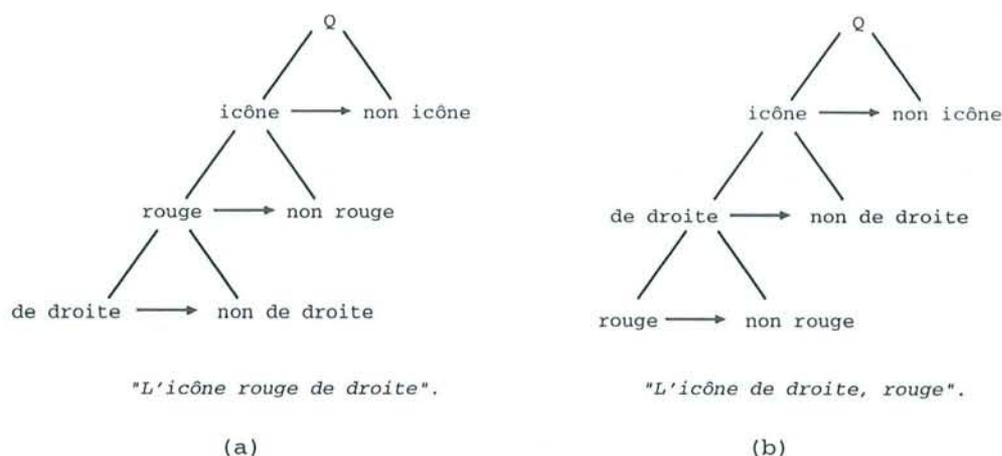


FIG. 3.6 – Schémas d'axiologie.

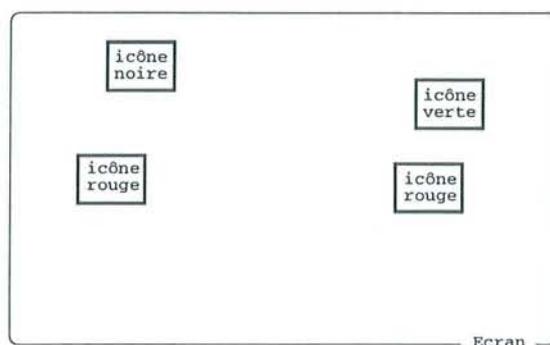


FIG. 3.7 – Le problème des propriétés non intrinsèques.

Une heuristique possible pourrait consister à appliquer les axiologies dans l'ordre suivant :

- appliquer la propriété de catégorie,
- appliquer les propriétés qui peuvent être évaluées de manière absolue,
- appliquer la (les) propriété(s) non intrinsèque(s)¹⁰.

Enfin, le modèle que nous venons d'exposer ne peut que décrire des groupes nominaux simples, en particulier, il n'est pas capable de décrire le fonctionnement de phrases du type "détruis l'icône qui se trouve à droite de la fenêtre". En effet, on voit très mal comment spécifier plus en avant l'opposition "icône" \rightarrow \neg "icône" à l'aide d'axiologies purement élémentaires car le modèle de Gaiffe et Romary n'est pas prévu pour prendre en compte des oppositions complexes du type "à droite de la fenêtre" \rightarrow \neg "à droite de la fenêtre".

du fonctionnement d'une phrase du type "détruis l'icône de droite" où l'axiologie "de droite" \rightarrow \neg "de droite" devrait faire appel au contexte pour discriminer l'icône considérée par rapport à des icônes qui seraient moins à droite ou pas à droite du tout.

10. Remarquons que le problème se complique sensiblement lorsque plusieurs propriétés non intrinsèques entrent en jeu. Il suffit de prendre l'exemple de "détruis la grande icône de droite" où seules des solutions *ad hoc* peuvent fournir des éléments de réponse sur l'ordre d'application des axiologies.

Au terme de cette étude, nous allons situer à présent notre problématique d'étude des références spatiales.

3.5 Problématique et intérêt de la référence spatiale

L'opération de référence spatiale est la faculté pour un être humain de retrouver la(les) bonne(s) entité(s) référencée(s) en faisant appel à tout type de processus cognitif d'origine spatiale. Les processus cognitifs peuvent regrouper des opérations mentales aussi diverses que la perception, la catégorisation, etc.

Le problème de la référence spatiale est essentiel dans un système de visualisation graphique car il concerne les trois opérations principales que l'on peut y faire :

- création d'objets : exemple : "*crée une icône à droite de la fenêtre rouge*".
- modification d'objets : exemple : "*agrandis la fenêtre verte de droite*".
- destruction d'objets : exemple : "*détruis l'icône de droite*".

Par ailleurs, il nous semble capital d'opposer un système opérant des références sur une scène d'objets à un système générant une description de ces objets : si pour un système descriptif il est nécessaire de faire une description précise des objets considérés (voir le système Soccer¹¹), en revanche, un système de référence peut se contenter du "strict minimum"¹², en tenant compte au maximum du contexte lui permettant de discriminer un (respectivement des) objet(s) par rapport à d'autres suivant l'une de ses (respectivement de leur) propriétés. Cette opposition entre un ou plusieurs objets suivant une ou plusieurs propriétés rejoint la notion d'axiologie que nous avons illustrée ci-dessus.

Pour illustrer la dualité référence/description, il suffit de prendre pour exemple la figure 3.8 pour laquelle, la phrase (79) serait amplement suffisante en référence alors que la phrase (80) pourrait être nécessaire pour faire un descriptif de ce qui se trouve affiché à l'écran :

Détruis l'icône qui se trouve au-dessus de la fenêtre. (79)

L'icône I1 est au-dessus de la fenêtre, légèrement à gauche et près du bord de l'écran. (80)

De notre point de vue, cette dualité référence/description trouve sa justification dans la notion de principe d'économie cognitive. En effet, quel que soit le contexte dans lequel se trouve plongé le système cognitif humain, ce dernier n'aura recours qu'à l'information qui lui est strictement nécessaire afin d'achever la tâche qui lui est impartie. Nous émettons l'hypothèse que la charge cognitive de cette tâche est généralement moindre en situation de référence qu'en situation de description.

11. Le système Soccer [Schirra, 1993] génère un commentaire d'un match de football à partir d'une séquence d'images.

12. À rapprocher de la notion de principe d'économie cognitive.

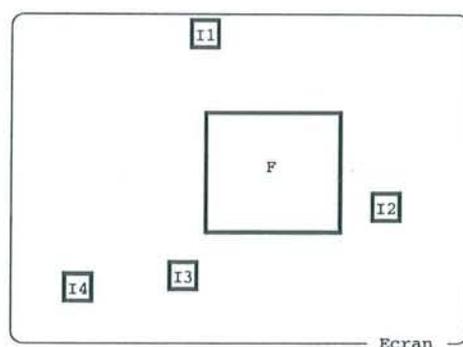


FIG. 3.8 – Avantages d'une optique différentielle.

3.5.1 Pourquoi s'intéresser aux références spatiales purement langagières ?

Il peut sembler curieux de s'intéresser à la référence spatiale purement langagière comme moyen de manipuler une interface graphique alors qu'une interface multimodale de type <geste + parole> par exemple semble à même d'offrir l'ensemble des fonctionnalités nécessaires tout en étant plus conviviale. Nous justifierons cependant notre étude par quatre raisons au moins qui nous apparaissent déterminantes :

- L'étude de type Magicien d'Oz de Mignot [1995] déjà mentionnée dans ce mémoire démontre que quand bien même l'utilisateur disposerait d'une interface multimodale, il ne se priverait en aucun cas d'effectuer des requêtes spatiales.
- Comme cela apparaît dans les thèses de Bellalem [1995] et Pouteau [1995], les actes purement gestuels se situent plus dans une perspective du "faire", alors que les requêtes du type <langage naturel> et <langage naturel + geste> relèvent plus du "faire-faire". Le second type de perspective est donc bien plus satisfaisant car l'utilisateur délègue davantage sa tâche à l'ordinateur.
- Une bonne interface homme-machine doit autoriser les phrases du type <requête spatiale>, <requête non spatiale> et <requête + geste> car on assiste à une très forte complémentarité entre ces trois possibilités de référence pour accéder à un univers d'objets. Voici respectivement un exemple de chacune de ces requêtes :

Détruis le carré de droite. (81)

Agrandis le triangle rouge. (82)

Peints ce rectangle en rouge. (+ geste) (83)

De notre point de vue, les références <requête non spatiale> prévalent lorsque le ou les éléments à référencer possèdent des caractéristiques intrinsèques que ne possèdent pas les autres éléments. Les références du type <requête spatiale> sont plus aisées lorsque le ou les éléments à référencer font partie d'une structure spatiale claire qui permet de l'(les) isoler par rapport aux autres éléments. Enfin, les références

de type <requête + geste> chevauchent les deux types précédents et apparaissent lorsqu'il y a la possibilité de faire naturellement un geste.

Pour illustrer ces trois types de référence, imaginons que l'utilisateur se trouve en présence de l'écran de la figure 3.9 et qu'il souhaite détruire respectivement l'objet 1, 2 et 3. Pour l'objet 1, une requête du type "détruis le cercle" (type <requête non spatiale>) nous semble plus naturelle par exemple que la requête "détruis le cercle qui se trouve en haut à gauche" qui est du type <requête spatiale>. En revanche, une requête comme "détruis cet objet" + geste ou "détruis ça" + geste semble naturelle.

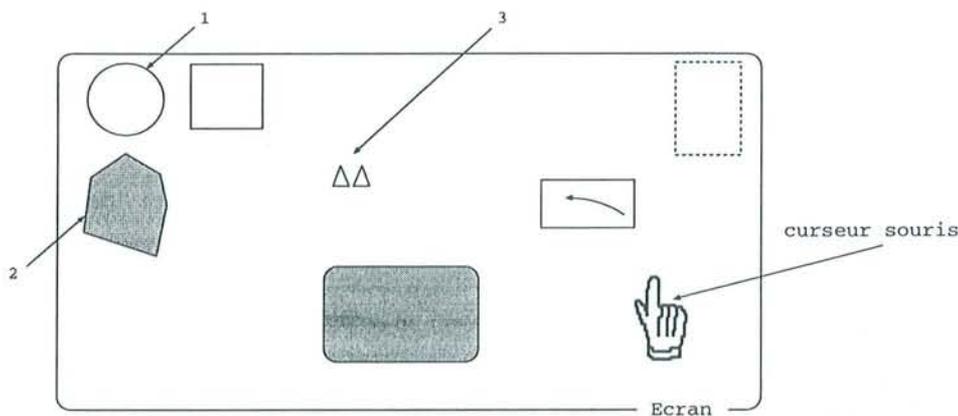


FIG. 3.9 – "Actes gestuels" versus "actes langagiers".

Pour la destruction du second objet, une requête purement spatiale simple¹³ semble difficile à mettre en œuvre car aucune structure spatiale ne se dégage clairement de la scène affichée (la requête "détruis l'objet qui se trouve en dessous du cercle" est possible mais longue à formuler). Une requête du type <requête non spatiale> semble également difficile à formuler du fait que l'objet 2 est difficile à nommer. Une requête du type <requête + geste> semble donc appropriée à faire référence à cet objet.

Enfin pour l'objet 3, produire un geste qui serait par exemple accompagné de la phrase "détruis ce triangle" n'est pas très aisé car il impose à l'utilisateur de faire un geste très précis compte tenu de la taille respective du triangle à détruire et de la taille du curseur souris. Une requête du type <requête non spatiale> n'est pas possible car aucune propriété intrinsèque ne permet de différencier les deux triangles. Une requête spatiale du type "détruis le triangle de gauche" nous semble donc parfaitement adaptée du fait de la forte structuration spatiale des deux triangles.

- Le point précédent semble démontrer qu'une bonne interface informatique doit autoriser des requêtes du type "détruis le triangle de droite". Dès lors, nous mettant à la place d'un utilisateur, nous serions quelque peu surpris si un tel système acceptait cette dernière phrase et refusait une requête semblable du type "détruis le triangle qui se trouve à droite du cercle". Ainsi, de proche en proche, il semble nécessaire

13. Par exemple pour l'objet en pointillés, la requête "détruis l'objet de droite" est une requête simple, alors que la requête "détruis l'objet qui se trouve en haut à droite" est déjà plus complexe.

d'autoriser toutes les requêtes spatiales purement langagières afin de ne pas tomber sur une inconsistance du système informatique qui pourrait dérouter l'utilisateur. Ce dernier point justifie donc *a posteriori* l'étude de phrases qui auraient pu sembler trop complexes au départ. Nous nuancerons cependant cette dernière remarque dans le sens où il y a bien entendu une frontière relativement floue au niveau de la complexité des phrases au-delà de laquelle l'utilisateur n'ira pas.

3.5.2 Notre contribution

a) Points généraux

De notre côté, nous allons nous intéresser au pilotage d'un système affichant des objets graphiques en deux dimensions en gérant en particulier les références spatiales qui peuvent porter sur ces objets. Pour simplifier, nous considérerons que ces objets peuvent être du type "cercle", "triangle", "carré" et "rectangle".

Par ailleurs, du fait de la multitude des problèmes à prendre en compte, nous avons été amenés à apporter les restrictions suivantes quant aux domaines d'application des modèles que nous allons présenter¹⁴ :

- Comme mentionné à la section précédente, nous ne considérons que le cas des références purement langagières, ce qui écarte les références multimodales du type <parole + geste>. Le lecteur intéressé par la prise en compte de la gestualité au sein d'un système de dialogue homme-machine finalisé pourra se reporter aux travaux de Bellalem [1993] et [1995]. Bellalem y décrit notamment que deux étapes sont identifiables dans le processus de traitement du geste. La première consiste en une analyse de la trajectoire gestuelle qui permet de la décomposer suivant des primitives représentant des formes simples (arc, segment de droite, points de rebroussement, points d'arrêt, etc.). La seconde est une étape d'interprétation qui s'appuie d'une part sur le résultat de l'étape précédente et d'autre part sur le contexte de production du geste. Ce contexte est formé du message langagier qui accompagne le geste, de l'état de la tâche et de la scène visualisée sur l'écran.
- Il n'y a pas toujours d'association biunivoque entre le nom commun utilisé dans l'application pour typer l'objet et le(s) nom(s) dont peut se servir l'utilisateur pour désigner ce même objet. Au contraire, ce nom procède souvent d'un point de vue adopté par l'utilisateur qu'il projette sur son application. Ainsi, si cette dernière donne *a priori* un type aux objets, l'utilisateur peut malgré tout utiliser un nom plus générique que celui associé au type de l'objet. Par exemple, imaginons une application qui serait capable de gérer l'affichage, la destruction, l'agrandissement, etc. de figures géométriques. Si un objet de cette application est étiqueté comme étant un "triangle", l'utilisateur peut quand même l'appeler une "figure". Autrement dit, il faut pouvoir détecter qu'un "triangle" est une "figure".

Pour gérer cette situation, il est nécessaire de répertorier au niveau du lexique l'ensemble des termes pouvant se rapporter à chaque type des objets de l'application.

14. Nous tenterons de lever certaines de ces restrictions dans le chapitre suivant.

Bellalem [1995] fait justement remarquer que le nom du type de l'objet, qui constitue le terme privilégié de l'application pour dénommer les objets, peut être considéré comme un prototype à partir duquel sont dérivés tous les autres termes susceptibles d'être employés par l'utilisateur.

Pour simplifier la suite de l'exposé, nous supposons que l'utilisateur n'emploie pas d'autre terme que celui qui correspond directement au type de chaque objet de l'application. Quatre types seront donc utilisables dans la suite de ce chapitre: le type "cercle", "triangle", "carré" et "rectangle".

- Les énoncés que nous considérerons doivent nécessairement faire intervenir une ou plusieurs références spatiales. Nous laissons donc de côté des requêtes non spatiales du type: "détruis le triangle rouge" qui ne font pas intervenir explicitement de préposition spatiale.

D'autre part, dans un souci de ne pas alourdir l'exposé, nous laisserons de côté les adjectifs pour nous focaliser exclusivement sur les aspects spatiaux. Comme on a pu le voir précédemment pour le modèle qui a été développé par Gaiffe et Romary, un adjectif ne réalise en fait qu'un filtrage supplémentaire (avant ou après le filtrage spatial) du ou des référents recherchés. À ce titre, il suffit de comparer la requête "détruis le triangle de droite" et "détruis le triangle de droite, rouge"¹⁵.

- Nous ne traiterons dans ce chapitre que des énoncés purement référentiels par opposition aux énoncés de positionnement. Ces derniers dont l'archétype se retrouve dans des phrases du type "Mets A prép. B" vont de notre point de vue plus loin que de simples énoncés référentiels. Pour illustrer ce dernier point, il suffit de prendre le cas de la phrase (84) qui nécessite d'une part de retrouver le "cercle" et "le triangle" et d'autre part de placer ce dernier au bon endroit. Le problème spécifique des énoncés de positionnement fera l'objet du chapitre 4.

Mets le triangle à droite du cercle. (84)

- Ce chapitre cherche surtout à apporter une solution aux problèmes touchant aux références spatiales qui s'effectuent directement à l'écran par opposition aux références anaphoriques. En effet, la référence anaphorique, quand elle existe (voir par exemple les phrases (85) et (86)) revient en fait à appliquer l'un des schémas que nous allons proposer au §3.7.4 sur une sous-partie de l'écran qui a déjà été mentionnée précédemment dans le discours. Nous laisserons donc de côté pour l'instant le cas de la référence anaphorique spatiale qui nous imposera un suivi discursif en ayant recours en particulier à la notion de cadre de référence.

Agrandis les carrés verts. (85)

Détruis celui de gauche. (86)

15. Bien entendu, il convient de porter une attention particulière sur l'ordre d'application des axiologies ainsi que nous l'avons relaté aux figures 3.6 et 3.7 (p. 87).

- Nous laisserons également de côté dans ce chapitre, le rôle joué par le prédicat dans le calcul des référents car si ce dernier est souvent déterminant dans le calcul référentiel, de notre côté, les phrases que nous nous proposons d’étudier ne comportent qu’un prédicat dont l’influence est relativement neutre sur le calcul du ou des référents comme par exemple : “*détruire*”, “*agrandir*”, etc.
- Une idée très répandue dans le domaine de la modélisation spatiale procède du fait que le site est généralement non ambigu, de position mieux connue que la cible et plus saillant, ce qui en fait l’entité idéale pour localiser une cible donnée (cf. le paragraphe 2.4). Nous verrons cependant à la fin de ce chapitre que la détermination du(des) site(s) n’est pas toujours aisée ce qui nous amènera à nuancer quelque peu “l’hégémonie localisatrice” du site. En effet, il semblerait que la cible puisse, dans une certaine mesure, contribuer à la détermination du site. Il suffit de prendre l’exemple de la figure 3.10 pour laquelle une requête du type “*détruis le triangle qui se trouve à droite du carré*”, qui présente dans l’absolu une ambiguïté sur le carré localisateur. Pour autant, le bon site est le carré C1 et non le carré C2 car le triangle T3 se trouve en fait en bas du carré C2. C’est donc le triangle T3 qui permet dans une certaine mesure de retrouver le carré C2. Cette dernière remarque semble montrer qu’il faille relativiser le rôle localisateur du site, la cible pouvant contribuer dans une certaine mesure à localiser le site et donc se localiser elle-même. Cependant, pour simplifier, nous supposerons dans ce chapitre que nous savons retrouver le site sans ambiguïté parmi les entités qui se trouvent affichées à l’écran. Dans un second temps, la cible sera retrouvée par extension progressive de la zone de recherche.

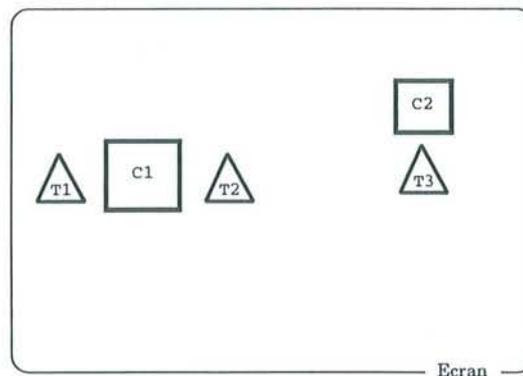


FIG. 3.10 – Quand le concept de cible empiète sur celui de site.

- Enfin, tout comme Hernández [1994], nous n’explicitons notre modèle que pour les prépositions “*à droite de*”, “*en haut à droite*”, “*à gauche de*”, etc. Nous dépasserons cependant l’approche de Hernández de deux manières. Tout d’abord, nous calculerons nos références “à” et “par rapport à” plusieurs objets, d’autre part, nous étudierons le cas des références absolues faisant intervenir les prépositions spatiales “*de droite*”, “*de gauche*”, etc.

b) Bilan des références spatiales à considérer

Voyons à présent, en tenant compte des différentes restrictions que nous avons envisagées dans le précédent paragraphe, les différentes références spatiales qu'il nous faut considérer pour piloter le système graphique.

b.1) La référence spatiale démonstrative

On peut distinguer la référence spatiale démonstrative anaphorique (phrase (88)) de la référence spatiale non anaphorique (phrase (89)). La première a été écartée à la précédente section car elle nécessite un suivi discursif. La seconde a été également écartée car elle s'appuie nécessairement sur un geste de désignation.

Agrandis les cercles de droite. (87)

Détruis celui de gauche. (88)

Détruis ce triangle. (+ geste) (89)

b.2) La référence spatiale indéfinie

Ce type de référence se fonde sur des constructions nominales du type “*un N*”, “*une N*”, “*des N*”, “*un des N*”, “*une des N*” dont voici quelques exemples :

Détruis un triangle à droite. (90)

Détruis l'un des cercles de droite. (91)

Détruis l'un des carrés qui se trouvent à gauche. (92)

Dans l'exemple de la phrase (90), le locuteur sous-entend par l'emploi du singulier, qu'il désire obtenir un seul triangle. Par l'emploi de l'indéfini, il sous-entend que l'espace désigné par la préposition “*à droite*” contient plusieurs éléments de la catégorie “*triangle*”. En effet, si cela n'avait pas été le cas, le locuteur aurait employé une expression définie et aurait donc dit : “*détruis le triangle de droite*”. En fait, le locuteur désire obtenir un triangle et un seul parmi ceux appartenant à la zone “*à droite*” mais n'importe lequel d'entre eux. Cet exemple démontre que le problème de la référence indéfinie est en fait très proche de la référence définie. Étant donné que les algorithmes que nous proposerons au §3.7.4 fonctionnent par extension de la zone de recherche jusqu'à isoler un objet unique ou un groupe d'objets unique, on peut astucieusement unifier le problème de la référence spatiale indéfinie avec celui de la référence spatiale définie. À titre d'illustration, la phrase “*détruis l'un des triangles de gauche*” énoncée à la figure 3.11 ne posera pas de problème car l'algorithme que nous proposerons au paragraphe 3.7.4 étend sa zone de recherche jusqu'à rencontrer un triangle, ici le triangle T4. De la même manière, les algorithmes que nous proposerons au paragraphe 3.7.4 seront capables de traiter les énoncés pluriels définis comme les énoncés pluriels indéfinis.

b.3) La référence spatiale définie

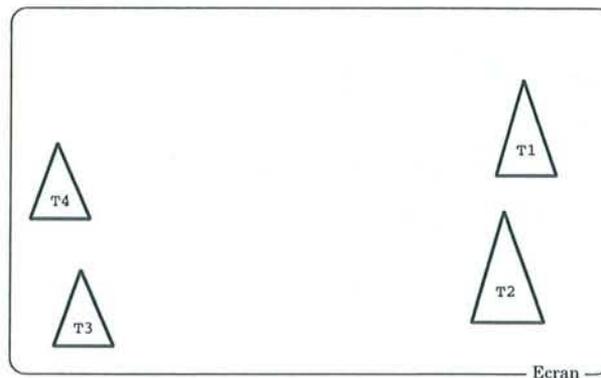


FIG. 3.11 – Unification de la référence définie et indéfinie.

Ce type de référence se fonde sur des constructions nominales du type “*le N*”, “*la N*”, “*l’ N*” et “*les N*”. De notre point de vue, parmi les références spatiales purement langagières, les références définies sont sans aucun doute celles qui sont les plus utilisées pour le problème qui nous intéresse car elles sont plus précises que les les références indéfinies précédemment étudiées. En voici quelques exemples :

Efface le cercle de droite. (93)

Agrandis le carré d’en bas. (94)

Détruis le triangle qui se trouve à droite du cercle. (95)

Efface les triangles. (96)

c) Bilan des phrases à considérer

Nous nous proposons de traiter les phrases possédant les structures syntaxiques suivantes :

<Verbe><Groupe Nominal><Prép. Spatiale> (97)

<Verbe><Groupe Nominal><Groupe Verbal>
<Prép. Spatiale> (98)

<Verbe><Groupe Nominal><Groupe Verbal>
<Prép. Spatiale><Groupe Nominal> (99)

Voici respectivement un exemple pour chacun des types de phrases considérés :

Détruis le triangle de droite. (100)

Détruis le triangle qui se trouve en haut à droite. (101)

Détruis le triangle qui se trouve en haut à droite du cercle. (102)

Nous supposons dans la suite que la phrase a été analysée à l'aide d'un analyseur syntaxique¹⁶ et qu'elle parvient à l'interpréteur décomposée sous la forme du 8-uplet suivant :

- (1) Verbe :
- (2) Déterminant de la(des) cible(s) :
 - genre : <féminin / masculin / neutre> ,
 - nombre : <singulier / pluriel> ,
 - type : <indéfini / défini / démonstratif>¹⁷
- (3) Catégorie de la(les) cible(s) :
- (4) Le séparateur :¹⁸
- (5) Déterminant du(des) site(s) :
 - genre : <féminin / masculin / neutre> ,
 - nombre : <singulier / pluriel> ,
 - type : <indéfini / défini / démonstratif> ,
- (6) Catégorie du(des) site(s) :
- (7) Préposition spatiale :
 - type de préposition spatiale : <simple / composée> ,
- (8) Type de localisation : <absolue / relative> .

Nous distinguons les prépositions spatiales simples du type “*de droite*”, “*à droite de*” des prépositions spatiales composées comme “*en haut à droite de*”¹⁹ à l'aide du champ “type de préposition”. Cette distinction sera utilisée par les algorithmes du paragraphe 3.7.4. Par ailleurs, les phrases qui comportent un site explicite comme “*détruis le carré qui se trouve à droite du rectangle*” seront qualifiées de relatives par opposition aux phrases du type “*détruis le triangle de droite*” qui porteront le qualificatif d'absolues.

Remarquons que tous les champs du 8-uplet ne parviennent pas toujours renseignés à l'interpréteur comme cela apparaît pour la représentation de la phrase (103) :

Détruis le triangle de droite. (103)

Détruis le triangle qui se trouve en haut à droite du cercle. (104)

Représentation de la phrase (103) :

- (1) Verbe : “*détruire*”,
- (2) Déterminant de la cible : “*le*”,
 - genre : masculin,
 - nombre : singulier,
 - type : défini,

16. Il pourrait s'agir par exemple d'un ATN (Augmented Transition Network). La théorie des ATN a été développée par Woods [1970].

17. Comme mentionné précédemment, notre étude porte sur le cas des références définies et indéfinies. Cependant, par souci de généralité pour la suite de nos travaux, l'analyseur syntaxique doit être capable d'analyser également des groupes nominaux composés de pronoms démonstratifs.

18. À la phrase (102), le séparateur est défini par le syntagme “*qui se trouve*”.

19. Le fait que nous parlions de préposition pour le syntagme “*en haut à droite*” est un abus de langage, le terme groupe prépositionnel serait sans doute plus adéquat.

- (3) Catégorie de la cible: “*triangle*”,
 - (4) Le séparateur: ×,
 - (5) Déterminant du site: ×,
 - genre: ×,
 - nombre: ×,
 - type: ×,
 - (6) Catégorie du site: ×,
 - (7) Préposition spatiale: “*de droite*”,
 - type de préposition spatiale: simple,
 - (8) Type de localisation: absolue.
- Représentation de la phrase (104):
- (1) Verbe: “*détruire*”,
 - (2) Déterminant de la cible: “*le*”,
 - genre: masculin,
 - nombre: singulier,
 - type: défini,
 - (3) Catégorie de la cible: “*triangle*”,
 - (4) Séparateur: “*qui se trouve*”,
 - (5) Déterminant du site: “*le*”,
 - genre: masculin,
 - nombre: singulier,
 - type: défini,
 - (6) Catégorie du site: “*cercle*”,
 - (7) Préposition spatiale: “*à droite de*”,
 - type de préposition spatiale: simple,
 - (8) Type de localisation: relative.

3.5.3 Conclusion

Nous retiendrons de cette étude la distinction qui a été faite entre un système opérant des références sur une scène d'objets et un système opérant une description de ces objets. Nous avons pu démontrer que cette distinction relevait du principe d'économie cognitive que nous avons défendu au premier chapitre. Nous avons également pu souligner que la cible avait parfois un rôle à jouer dans les phénomènes de localisation. Cette remarque essentielle sera corroborée pour d'autres exemples dans la suite de cette thèse.

Par ailleurs, nous avons démontré l'utilité d'une étude des références spatiales purement langagières qui vient en complémentarité d'une étude des références purement gestuelles et d'une étude des références du type <parole + geste>.

Enfin, la problématique de la suite de ce chapitre portera sur l'étude des expressions définies et indéfinies purement spatiales.

3.6 Une étude détaillée de la préposition “à gauche de”

Cette partie fera l’objet de l’étude de plusieurs modèles qui pourraient s’adapter au problème de la référenciation d’objets d’un espace de visualisation graphique.

3.6.1 Un modèle discrétisant l’espace

Si nous assimilons les quatre types d’objets que nous considérons (les carrés, les rectangles, les cercles et les triangles) à leurs enveloppes convexes rectangulaires, nous pouvons reprendre l’étude de Mukerjee [1989]²⁰. L’auteur y adapte la logique de Allen²¹ pour des dimensions supérieures à 1. Il projette chaque solide sur chacun des trois axes et calcule alors, selon le modèle de Allen, les positions respectives des segments (ou intervalles) résultants sur ces axes.

Une première difficulté sous-jacente au modèle de Mukerjee provient du fait que tout objet qui se trouve dans la zone grisée de la figure 3.12.(a) se trouve exactement à la même position par rapport au carré C.

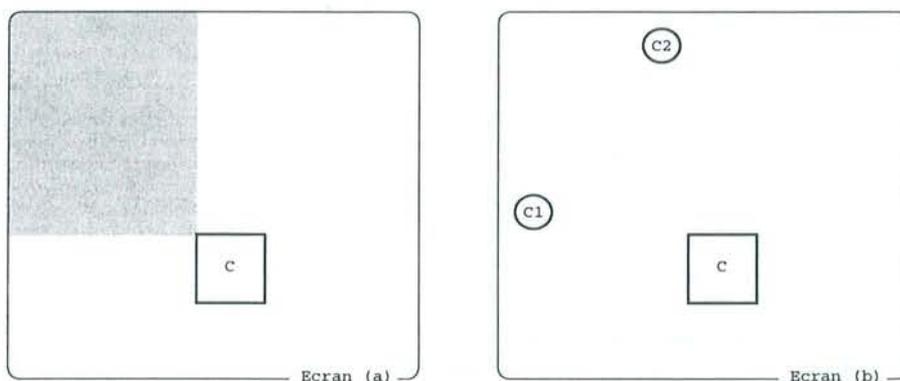


FIG. 3.12 – Un modèle imprécis.

Cependant, en reprenant le diagramme de [Briffault, 1992] p. 70 (voir figure 3.13²²), nous pensons au contraire que le cercle C1 de la figure 3.12.(b) pourrait être référencé sans aucun problème comme étant “le cercle qui se trouve à gauche du carré”. Le cercle C2 pouvant quant à lui être référencé comme étant “le cercle qui se trouve au-dessus du carré”. Ce petit exemple illustre le caractère quelque peu trop grossier du modèle de Mukerjee

20. L’étude de [Mukerjee et Joe, 1990] plus récente ne peut pas s’appliquer ici car elle ne fonctionne qu’avec des parallélépipèdes ayant des directions frontales non parallèles.

21. Allen [1983] et [1984] a développé une théorie temporelle qui s’appuie explicitement sur des intervalles en ne représentant que l’information effectivement connue du système. Allen introduit une série de treize relations (*before, equal, meets, etc.*) pouvant lier deux intervalles de manière non exclusive (plusieurs relations peuvent apparaître entre deux intervalles). Le temps n’est donc défini que de manière relative dans ce modèle, à l’exclusion de toute référence à des bornes numériques. Un intervalle particulier n’est défini qu’à partir des seules relations qu’il établit avec ses voisins, formant de cette manière un réseau de représentation au sein duquel apparaissent certaines contraintes.

22. Les zones hachurées en traits épais correspondent aux zones d’acceptabilité maximale de la préposition, tandis que les secteurs hachurés en clair correspondent aux zones où l’acceptabilité de cette préposition est supérieure à celle des autres.

pour le problème qui nous intéresse. Il semble donc que la discrétisation de l'espace en termes d'intervalles de ce type ne soit pas assez fine.

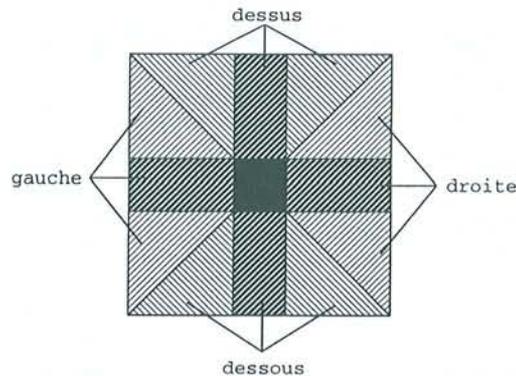


FIG. 3.13 – Acceptabilité des prépositions selon Briffault.

Plus généralement, on pourra remarquer que les difficultés rencontrées par ce modèle proviennent du fait qu’il se centre presque exclusivement sur le site. Pour autant, il est bien clair à la figure 3.12 que la cible a également un rôle à jouer dans le choix de la préposition spatiale qui décrit sa position par rapport au site.

Enfin, le principal reproche que nous ferons à ce modèle provient de sa perte de plausibilité cognitive que possédait l’approche de Allen dans le domaine temporel. En effet, nous autres humains, ne passons pas notre temps à décomposer le monde suivant deux ou trois axes pour déterminer ensuite l’ensemble des relations qui sont mises en jeu entre les entités du monde concernées.

3.6.2 Un modèle linguistique

L’approche de Vandeloise, [Vandeloise, 1986], qui a été exposée en détails au chapitre 2 se heurte également à quelques difficultés. L’auteur propose la règle d’usage suivante pour les prépositions “à droite de”/“à gauche de” :

G/D: A est “à droite”/“à gauche de” B si A se trouve du côté “droit”/“gauche” de l’orientation latérale²³ de B.

Suivant cette approche, le cercle C2 de la figure 3.14 serait autant à gauche du carré C que le cercle C1, ce qui bien entendu est trop grossier et pratiquement faux dans une optique différentielle. Tout comme au paragraphe précédent, C2 serait vu comme “au-dessus du cercle” plutôt que “à gauche du cercle”. Vandeloise sent cependant qu’il y a un problème. Ainsi, pour la préposition “au-dessus de”, l’auteur indique que la relation entre les deux objets doit être saillante [Vandeloise, 1986] (p. 104). On pourra remarquer que

23. **Orientation latérale** : il s’agit d’une ressemblance de famille dont les principaux traits sont la *direction latérale* et la perpendiculaire à l’*orientation générale*.

Orientation générale : ressemblance de famille qui compte pour principaux traits : la *direction frontale*, la *direction du mouvement*, la *ligne du regard*, la *direction dans laquelle sont dirigés les autres organes* : la *perception*, l’*odorat*, l’*ouïe*... (voir le chapitre 2).

le fait de parler de saillance ici revient en fait à imputer à la cible un rôle plus important que celui dont elle jouit habituellement.

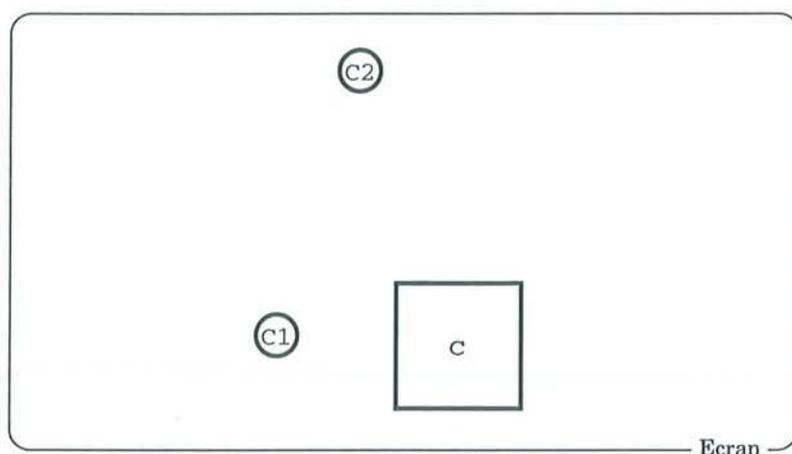


FIG. 3.14 – Différentes acceptabilités de “à gauche de”.

3.6.3 Un modèle à base de nuages d’acceptabilité

Le système SOCCER ([Wahlster, 1987; Schirra, 1993]) est capable de générer un compte rendu simultané d’événements à partir d’une séquence d’images en cours de traitement, l’application étant celle de commentaires en direct de matchs de football. Pour réaliser ceci, Schirra s’appuie, entre autres auteurs, sur des nuages de probabilité²⁴, dont la forme est définie pour chaque préposition et s’instancie suivant le contexte (voir par exemple à la figure 3.15.(a) un nuage de probabilité possible pour la préposition “à gauche de”).

Si l’on tente d’appliquer tel quel le système de Schirra à nos objets (en les restreignant à leur centre de gravité comme l’auteur le fait pour les joueurs de football), cela ne fonctionne pas (voir figure 3.16) où l’on a envie de dire : “*R1 est à gauche de C*” et “*R2 est au-dessus de C*” et pas quelque chose du type : “*R1 est en haut à gauche de C*”?! et “*R2 est en haut à gauche de C*”?! du fait de la position relative de leurs centres de gravité G1 et G2.

Là encore, on s’aperçoit que les difficultés rencontrées proviennent finalement du fait qu’il faut obtenir un modèle où le site et la cible sont globalement confrontés et pas un modèle où seul le site contraint la forme du nuage.

La solution de prendre les quatre coins n’est pas meilleure car ces derniers peuvent se trouver hors du nuage de probabilité que définit l’entité localisatrice. Il suffit de prendre l’exemple de la figure 3.15 sur laquelle le rectangle R1 se trouve bien à gauche du petit carré C, quand bien même ses coins ne font pas partie du nuage de probabilité issu du carré C. La solution que nous préconisons dans ce cas, est la prise en compte simultanée des 4 coins et du centre de gravité de l’objet référencé. Il suffit alors de chercher quel est

24. Nuages de probabilité : matérialisent le degré d’applicabilité (compris dans l’intervalle [0;1]) de la préposition concernée.

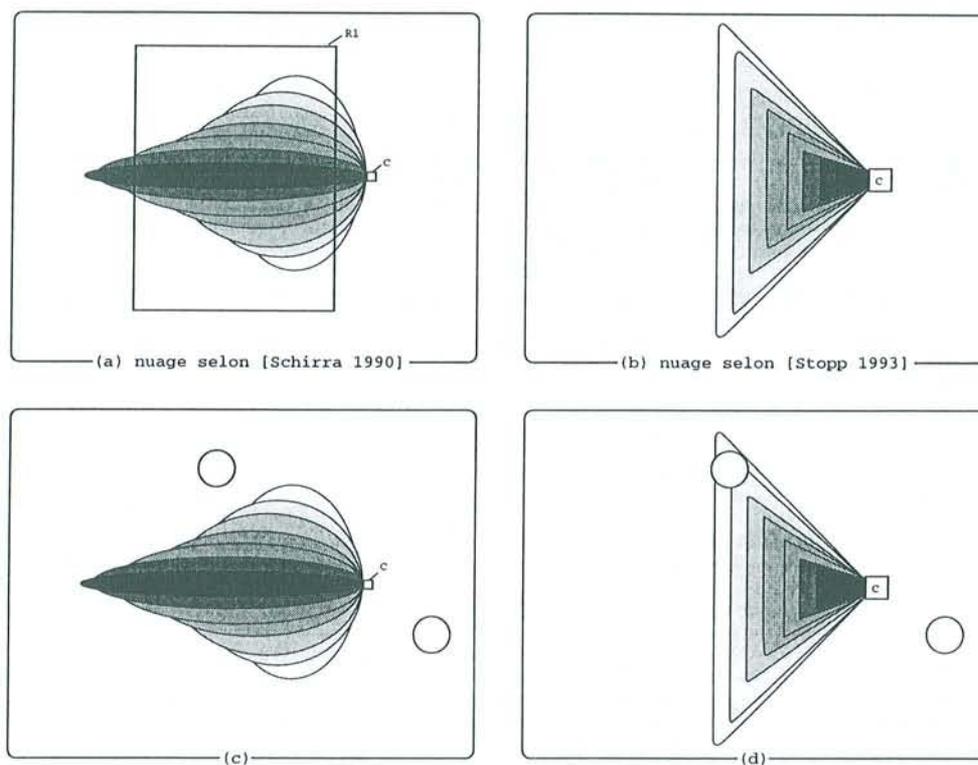


FIG. 3.15 – Deux nuages de probabilité pour la préposition “à gauche de”.

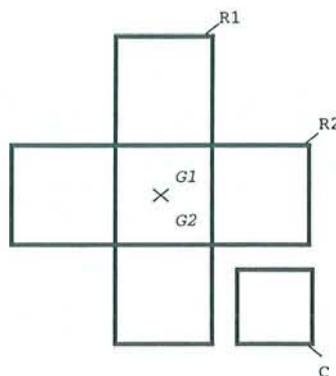


FIG. 3.16 – Imprécision des centres de gravité.

le nuage (“à gauche de”, “au-dessus de”, etc) qui maximise la somme des densités des quatre coins et du centre de gravité. Cependant, il se peut, là encore, qu’en cas de grande différence de taille entre C et R1, le centre de gravité de R1 ne fasse, lui non plus, pas partie du nuage. Il suffit alors de calculer la position de C par rapport à R1 et de prendre la préposition converse.

Par ailleurs, on peut débattre du choix à porter sur les nuages de probabilité: par exemple, vaut-il mieux choisir celui de Schirra ou celui de Stopp (figures 3.15.(a) et 3.15.(b)) ? Nous plaiderons en faveur du nuage de probabilité de Stopp car ce dernier

semble plus exact si on le rapproche des zones d'acceptabilité définies par Briffault (cf. figure 3.13, p. 99). En effet, on voit mal pourquoi limiter autant la portée verticale du nuage lorsque l'on s'éloigne du site. Voir à ce propos les figures 3.15.(c) et 3.15.(d) pour une requête du type "détruis le cercle qui se trouve à gauche du carré" où le cercle ne pourra être correctement localisé à la figure 3.15.(c).

3.6.4 Un modèle fondé sur des inclusions de portions de surface

Wazinski [1991] et [1993] travaillant en 2D, étudie précisément la position relative de deux rectangles. Voulant référencer un rectangle R2 par rapport à un rectangle R1, Wazinski découpe l'espace 2D en 9 sous-espaces définis à partir de R1 (voir figure 3.17.(a)). Il calcule ensuite la proportion de surface de R2 qui se trouve dans tel ou tel secteur.

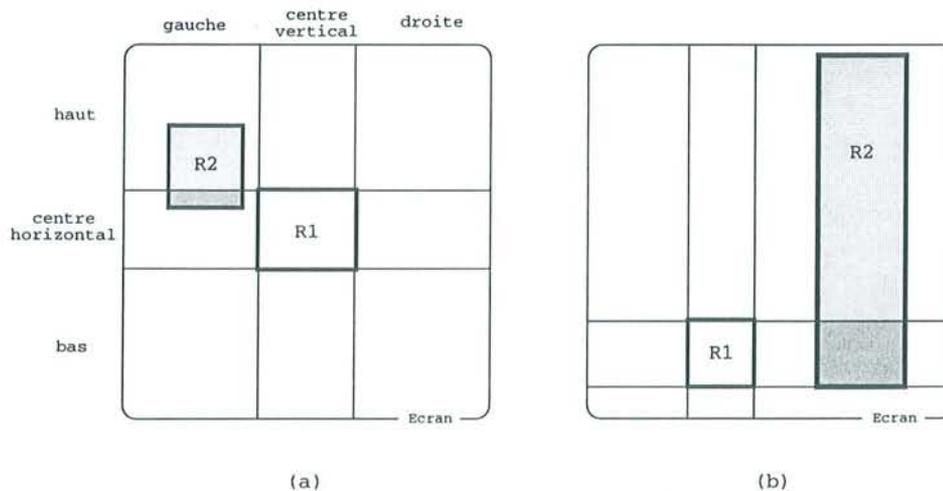


FIG. 3.17 – L'approche de Wazinski, ses limites.

Ce modèle lui permet de prendre en compte les trois types de localisations qu'il distingue :

- les localisations relatives : "détruis l'objet qui se trouve à droite de l'objet X".
- les localisations absolues : "détruis l'objet de droite".
- les localisations dans les coins : "détruis l'objet qui se trouve dans le coin inférieur droit de l'écran".

L'idée semble séduisante car comment approcher plus finement la position d'un rectangle, par rapport à un autre, qu'en prenant en compte la portion de sa surface qui se trouve incluse dans telle ou telle zone de l'espace? Il semble cependant que ce type d'approche présente des limites. Pour s'en convaincre, il suffit de prendre pour exemple la figure 3.17.(b) où le modèle de Wazinski conclurait que R2 se situe en haut à droite de R1 car la surface qui se trouve dans le secteur en haut à droite de R1 est bien plus importante que celle se trouvant dans le secteur directement à droite de R1.

Là encore, le modèle proposé présente certaines lacunes qui proviennent du fait que Wazinski a omis d'analyser les rôles respectifs de la cible et du site dans les paramètres de son modèle.

3.6.5 Un modèle à base de zones d’acceptabilité

Briffault [1992], [1993] esquisse ce que l’on pourrait dénommer une vision cognitive des prépositions “à droite de, à gauche de, au-dessus de, en dessous de” (voir figure 3.13 pour rappel). Adoptant ce système à nos objets graphiques, l’algorithme qui consisterait à tester si le centre de gravité appartient à tel ou tel secteur ne fonctionne bien entendu pas pour la même raison que celle explicitée à la figure 3.16. Néanmoins, le système fonctionne bien si l’entité à localiser appartient totalement à l’un des secteurs. L’auteur ne dit cependant rien lorsqu’un rectangle chevauche plusieurs secteurs. Notre propos, au paragraphe 3.7.4, sera justement d’étendre ce modèle lorsqu’un rectangle chevauche plusieurs secteurs d’acceptabilité.

3.6.6 Le modèle de Hernández

Le modèle de Hernández [1994] a été exposé en détail au paragraphe 2.6. À la différence des précédents modèles, l’approche de Hernández présente l’avantage de pouvoir rechercher plus finement la cible en choisissant le système d’axes (parmi les trois possibles) qui est le plus approprié à discriminer la cible recherchée. Ainsi, ce modèle est capable de retrouver le bon cercle à la figure 3.12.(b) pour une requête du type “*détruis le cercle qui se trouve à gauche du carré*”²⁵. Les modèles de Mukerjee, Wazinski et Vandeloise n’en seraient pas capables car ils placeraient les cercles C1 et C2 dans la même zone, relativement au carré C, à la figure 3.12.(b). Le modèle de Schirra y parviendrait en revanche car les cercles C1 et C2 n’occuperaient pas la même position dans le nuage de probabilité défini relativement au carré C.

Finalement, le reproche que nous ferons à cette approche tient précisément dans l’éclatement en trois niveaux que propose Hernández. L’unification de ces niveaux, tout en préservant son aptitude à retrouver précisément la cible, fera partie de nos préoccupations au paragraphe 3.7.

3.6.7 Autres limites de ces modèles

Nous avons laissé pour l’instant de côté le problème de la référence absolue, celui de la référence plurielle et enfin celui de la portée localisatrice limitée du site. La référence absolue permet de localiser un objet sans que le site ne soit explicitement donné (cette localisation opère donc très souvent sur la base de tout l’écran)²⁶ :

Détruis le carré de gauche. (105)

Si les modèles de Mukerjee, Briffault, Vandeloise et Schirra n’ont pas été explicitement prévus pour le cas de la localisation absolue, celui de Wazinski semble y parvenir en

25. Pour la figure 3.12.(a), le système d’axes de niveau 1 suffirait tandis qu’à la figure 3.12.(b), le système d’axes de niveau 2 suffirait.

26. M. Aurnague [1991] parlerait de localisation interne. En effet la préposition spatiale va permettre de localiser la position de la cible relativement à une portion de l’écran.

faisant jouer au centre de l'écran le rôle de site implicite²⁷. Cependant, certaines difficultés subsistent, il suffit pour s'en convaincre d'étudier le cas de la figure 3.18.(a) sur laquelle l'utilisateur effectue la requête "détruis le carré de droite"²⁸. Le modèle de Wazinski ne parvient pas à isoler le carré C2 car ce carré ne se situe pas explicitement à droite de l'écran. De notre point de vue, ces difficultés proviennent sans doute de la forte opposition spatiale qui existe entre les carrés C1 et C2 et que ne parvient pas à prendre en compte un modèle qui serait centré sur l'écran.

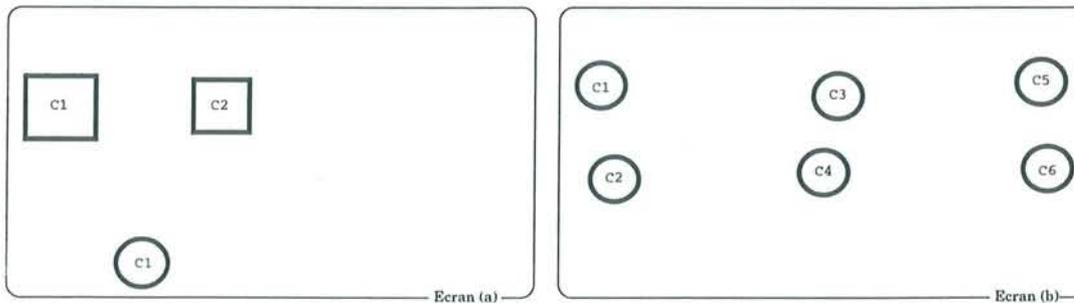


FIG. 3.18 – Les problèmes de la référence absolue et plurielle.

Le cas de la référence plurielle (qui met en jeu la localisation de plusieurs cibles) semble encore plus délicat. Pour s'en convaincre, étudions la figure 3.18.(b) pour la requête "détruis les cercles de droite" en considérant que le problème de la référence absolue est résolu pour les références au singulier. Pour pouvoir retrouver correctement les cercles C5 et C6, il est nécessaire de repérer la structuration perceptive de la scène affichée où se détachent clairement trois groupes homogènes perceptifs : celui constitué par les cercles C1 et C2, par les cercles C3 et C4, et finalement par les cercles C5 et C6. Aucun des modèles que nous avons exposé n'est prévu pour ce cas de figure.

Nous terminerons cette étude critique des modèles existants par le problème de la portée localisatrice limitée du site. Ainsi, prenons le cas de la figure 3.19 pour une requête du type : "détruis le cercle qui se trouve à droite du carré". Excepté le modèle de Stopp [1993], aucun des modèles que nous avons vus ne parviendrait à proposer uniquement le cercle C2 et non les deux cercles C1 et C2 car la zone d'acceptabilité recouvre à chaque fois ces deux objets. Le nuage de probabilité tel que l'avait proposé Schirra [1990] (figure 3.15 p. 101) échouerait également dans la distinction entre les deux cercles C1 et C2 car ils font partie du même nuage de probabilité. Si le nuage de probabilité de Stopp paraît idéal, un risque demeure cependant. On peut en effet imaginer certains contextes marqués où la distance séparant la cible recherchée du site est très grande. La cible pourrait alors se trouver en dehors du nuage de probabilité, rendant sa localisation impossible.

Une solution possible pourrait consister à étendre progressivement l'aire de recherche à partir du site sans lui donner une extension *a priori* qui serait entièrement et uniquement

27. Wazinski propose que le site occupe le tiers de l'écran au centre.

28. La requête : "détruis le carré du milieu" est également tout à fait plausible.

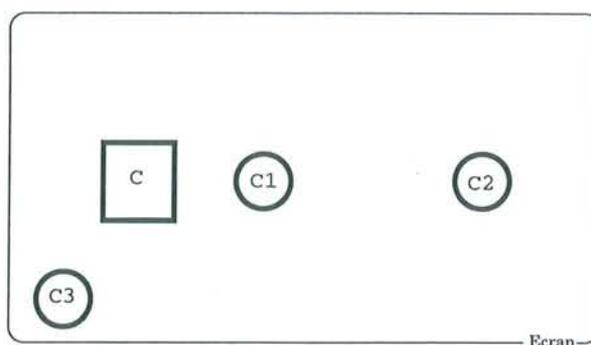


FIG. 3.19 – *Le problème de la portée localisatrice limitée du site.*

fondée sur le site. Nous exploiterons cette solution à base d'extension de l'aire de recherche pour l'algorithme que nous proposerons au §3.7.4.

3.6.8 Conclusion

Différents points ressortent de cette étude critique. Tout d'abord, pratiquement tous les modèles que nous venons d'étudier accordent un quasi monopole au site, ce qui nuit bien souvent à une recherche fine de la cible. Nous avons également relaté le fait que les modèles étudiés laissent de côté le problème de la référence absolue, excepté le modèle de Wazinski qui pour autant n'apportait pas une solution satisfaisante. D'autre part, une étude succincte de la référence plurielle démontre la nécessité d'aller plus loin qu'un simple découpage de l'écran en plusieurs zones. Au contraire, l'adjonction d'une composante perceptive permettant de structurer la scène affichée s'avère nécessaire. Enfin, nous avons pu démontrer que tous les modèles abordés, excepté celui de Stopp (sauf à de rares exceptions), étaient inaptes à tenir compte de la portée localisatrice du site. Ces différentes remarques nous ont incité à élaborer notre propre modèle que nous présentons au prochain paragraphe.

3.7 Vers une autre approche

3.7.1 Introduction

Nous détaillerons dans ce paragraphe l'évolution de notre réflexion concernant l'élaboration d'un modèle de calcul des référents spatiaux dans un système affichant des objets graphiques. Cette progression apparaîtra tout d'abord dans la réflexion que nous avons menée sur le rôle possible de la notion de saillance et celui de la notion de prototype. Les absences que nous y relèverons nous inciteront à nous tourner vers un autre type d'approche.

Il est important de noter que du fait de la position canonique de l'utilisateur face à son écran, de la nature des entités qu'il manipule et le fait qu'elles ne se chevauchent

pas, la seule orientation que nous rencontrons est l'orientation déictique²⁹. Pour un exposé détaillé concernant les conflits possibles entre les différentes orientations, voir [Sondheimer, 1976], [Retz-Schmidt, 1988], [Garnham, 1989], [Carlson-Radvansky et Irwin, 1993].

3.7.2 Rôle possible de la notion de saillance

Si l'utilisateur souhaite désigner un ou plusieurs objets, on peut penser qu'il le fera prioritairement à l'aide de ses(leurs) propriétés saillantes, c'est-à-dire que la référence se fera par discrimination sur une ou plusieurs propriétés que possède(nt) l'(les) entité(s) référencée(s) et que ne possèdent pas les autres. Ces propriétés saillantes ne se fondent pas que sur des propriétés spatiales. En effet, selon le principe de pertinence, s'il est plus facile pour l'utilisateur de faire référence à un objet à l'aide d'une propriété non spatiale (qui peut par exemple faire appel à une propriété intrinsèque de l'objet) il le fera.

La seule difficulté consiste alors à construire cet ensemble de propriétés saillantes qui fournira une description possible (qui n'est pas unique) de ce qui se trouve affiché à l'écran. Nous n'entrerons pas dans le détail de la construction de ces différentes propriétés car cette solution présente l'immense inconvénient de nécessiter un recalcul complet de l'ensemble des propriétés saillantes dès qu'il y a ajout ou suppression d'un ou plusieurs objets à l'écran. Pour cette raison, nous ne développerons pas plus ce type de modélisation.

Dans le type d'approche que nous venons d'esquisser, l'ordinateur part de l'univers d'objets pour lequel il essaie de donner une représentation. Il essaie ensuite de faire correspondre directement l'(les) objet(s) que l'utilisateur souhaiterait désigner dans sa phrase avec la représentation qu'il a de ces objets. Réciproquement, on peut adopter la démarche inverse en partant de la phrase énoncée et rechercher l'(les) entité(s) référencée(s) en se tournant vers la théorie du prototype.

3.7.3 Rôle possible de la notion de prototype

Au vu des nombreux travaux réalisés sur la théorie du prototype, nous avons tenté de réaliser un modèle de calcul des référents se fondant sur ce concept.

Pour mémoire, la notion de prototype a été introduite pour remettre en cause la vision classique de la catégorisation (qui remonte à Aristote), aussi connue sous le nom de *modèle* ou *théorie des conditions nécessaires et suffisantes*. On considérait, avant les travaux de Rosch [1978] que pour qu'un objet appartienne à une catégorie, il fallait que cet objet satisfasse à un certain nombre de conditions nécessaires et suffisantes. La théorie du prototype de Rosch substitue à cette vision rigide de la catégorisation une vision beaucoup plus souple, selon laquelle un objet appartient ou n'appartient pas à une catégorie selon le degré de ressemblance qu'il entretient avec un exemplaire particulier de

29. Pour mémoire, on définit classiquement ces trois orientations de la manière suivante :

- Orientation déictique : orientation fondée sur le locuteur.
- Orientation extrinsèque : orientation à partir des traits marquants de l'univers comme la gravité.
- Orientation intrinsèque : orientation à partir d'un objet.

cette catégorie ou *prototype*.

Voyons à présent comment la notion de prototype pourrait avoir un rôle à jouer dans le calcul du(des) référent(s) spatiaux recherché(s). Intéressons nous tout d'abord à des requêtes de l'utilisateur qui ne mettent en cause qu'un seul objet, c'est-à-dire dont la référence spatiale ne concerne qu'un seul objet de la tâche. Dans une tentative d'implémentation informatique, une solution pourrait consister à construire à partir de l'expression linguistique un prototype \mathcal{P} de l'objet référencé, c'est-à-dire une entité abstraite qui représente le plus fidèlement possible les attributs de l'objet que l'utilisateur souhaite désigner (ces attributs sont constitués du type de l'objet, de sa position, de sa taille, de sa couleur, etc.). Dans un second temps, il suffit de rechercher l'objet \mathcal{O}_i qui s'écarte le moins possible de ce prototype, c'est-à-dire celui qui vérifie au mieux :

$$\text{Min}_{i=1\dots n} \left(\sum_{j=1}^m \text{divergence}_j(\mathcal{P}, \mathcal{O}_i) \right)$$

Ainsi, pour la requête "*détruis le carré de gauche*" qui serait énoncée à la figure 3.20.(a), il suffit dans un premier temps de construire un prototype d'un carré que l'on situerait sur la gauche de l'écran (voir la position approximative du prototype en pointillés à la figure 3.20.(b)). Dans un second temps, on recherche l'objet qui présente le moins de divergences avec le prototype. Cet objet qui est déterminé sans ambiguïté à la figure 3.20.(b) n'est autre que le carré C1.

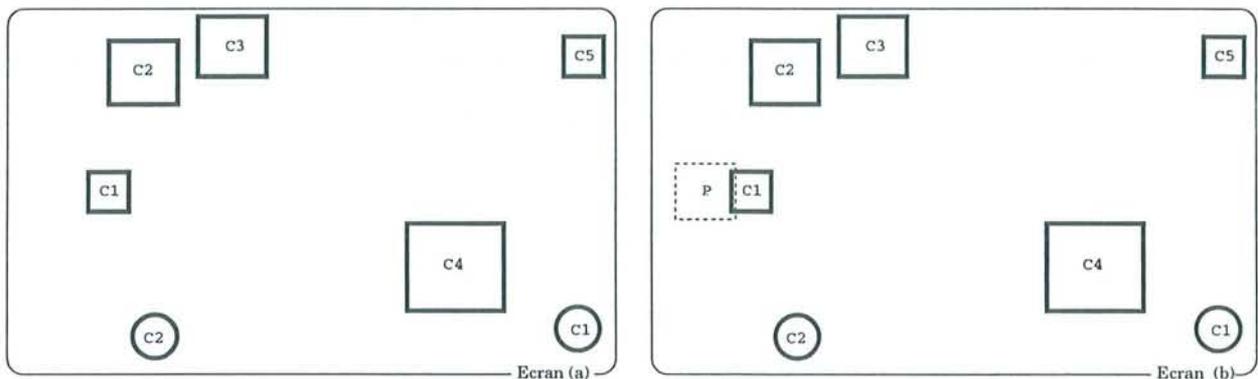


FIG. 3.20 – De la difficulté à construire le prototype.

Le point délicat de cet algorithme réside précisément dans la construction du prototype. Ainsi, prenons la figure 3.20.(a) pour laquelle l'utilisateur aurait énoncé la phrase (106) :

Détruis le carré de droite. (106)

Si l'on sait où construire le prototype (approximativement à droite!), on ne sait quelle taille (longueur, largeur) lui donner. Suivant cette taille, on risque alors d'obtenir une

grande valeur pour la $divergence_{taille}(\mathcal{P}, C5)$ et de rejeter C5 au profit de C4?! Sans entrer dans le détail du choix des propriétés de divergence à considérer³⁰, l'exemple précédent démontre qu'il semble impossible de construire un prototype complètement *a priori* puis de le confronter ensuite aux objets qui se trouvent affichés sur l'écran. Il semble au contraire qu'il faille construire, au départ, un prototype qui ne se compose que d'attributs très pauvres et généraux (par exemple, dans le cas de la recherche d'un rectangle qu'il faudrait détruire, on cherchera en priorité un objet de nature rectangulaire sur l'écran et pas un cercle par exemple). Dans un second temps, ces attributs vont s'instancier au fur et à mesure, suivant la scène affichée, pour isoler l'unique objet qui a été désigné.

Sans aller plus en avant dans ce type de méthodologie, nous allons voir à présent que la prise en compte de la scène affichée, en vue de construire le prototype, est parfois très complexe à réaliser. Prenons l'exemple de la figure 3.21.(a) pour laquelle on souhaite à nouveau détruire le carré de droite. Quand bien même nous aurions réussi à construire un prototype qui soit relativement acceptable (voir le carré P en pointillés à la figure 3.21.(b)), le risque est grand sur cette figure de choisir au dernier moment le carré C4 plutôt que le carré C5 pour des raisons de proximité manifeste. La solution pourrait consister alors à construire une relation de divergence par préposition spatiale (par exemple pour la préposition "à droite de", le critère déterminant de divergence serait la distance horizontale entre le prototype et l'objet \mathcal{O}_i considéré). Outre le foisonnement des relations de divergence à considérer³¹, il semble que l'on perdrait définitivement le caractère général de la théorie du prototype en ayant recours à de telles restrictions.

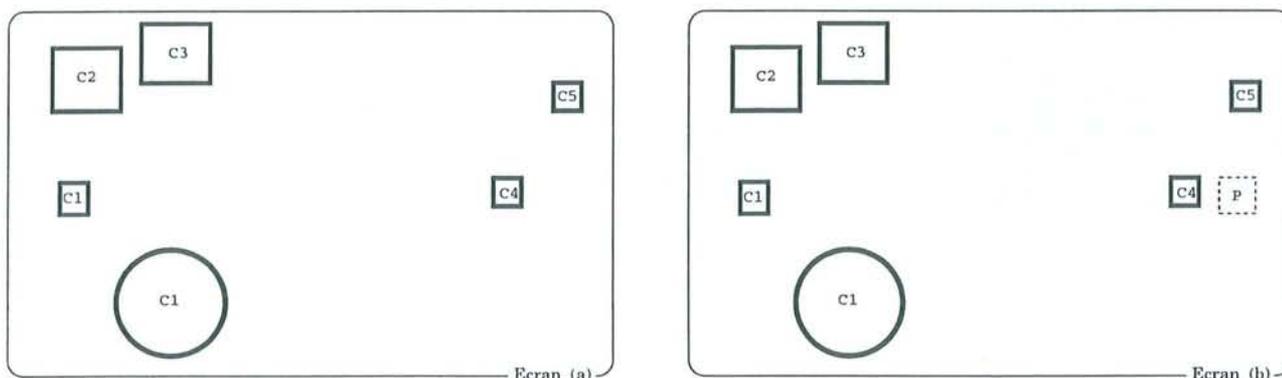


FIG. 3.21 – Les errements du prototype.

Notons enfin que nous avons laissé de côté jusqu'ici le cas de la référence à plusieurs objets (voir par exemple le cas de la figure 3.22 pour une requête du type :

30. Ce point est assez délicat : il suffit de prendre comme exemple le cas des propriétés de couleur qui entreraient en jeu si l'utilisateur disait par exemple "détruis le carré vert". Dans ce dernier cas, il faudrait nécessairement que $divergence_{couleur}(\mathcal{P}, Carre_{rouge}) \gg divergence_{taille}(\mathcal{P}, Carre_i)$ pour ne pas proposer à l'utilisateur un carré qui ne serait pas vert. Cet exemple semble démontrer qu'on ne peut pas placer les propriétés de couleur au même niveau que les propriétés de position et de taille.

31. Qui ne résoudraient pas pour autant des exemples du type : "détruis le grand carré de droite" où deux relations de divergence sont à prendre en compte ce qui ne nous garantit pas le bon résultat suivant les autres objets affichés.

“détruis les carrés de droite”) car il devient dans ce cas délicat de fixer une borne pour $\text{Min}_{i=1..n}(\sum_{j=1}^m \text{divergence}_j(\mathcal{P}, \mathcal{O}_i))$ en deçà de laquelle un carré serait dit à droite, et au-delà où il ne le serait plus.

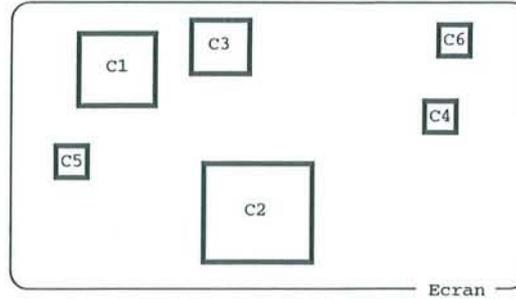


FIG. 3.22 – Le cas de la référence à plusieurs objets.

Les carences de ce modèle nous ont incité à nous tourner vers un autre modèle comme nous allons le voir tout de suite.

3.7.4 Vers un modèle de calcul des référents spatiaux

L'idée qui sous-tend le modèle que nous allons présenter à présent consiste à étendre progressivement une aire de recherche jusqu'à isoler un objet ou un groupe homogène perceptif (pour le cas où plusieurs objets sont concernés). Il est à noter que l'extension de l'aire spatiale de recherche dépend de la préposition spatiale qui est mise en jeu.

Les algorithmes que nous allons exposer ci-dessous ont été répartis tout d'abord en deux classes selon que l'on cherche à localiser une entité ou un groupe d'entités. Dans un second temps, chaque classe est décomposée en deux sous-classes selon que nous sommes en présence d'une localisation absolue ou relative. Enfin ces deux sous-classes se redécoupent à nouveau en fonction de la nature de la préposition spatiale qui est mise en jeu : cette dernière est soit simple soit composée.

Remarque : les algorithmes présentés ici concernent les prépositions “à droite de”, “de droite”, “en haut à droite” et sont bien sûr extensibles aux autres prépositions projectives (“à gauche de”, “au-dessus de”, “en dessous de”, “de gauche”, etc.) par symétrie.

a) Localisations d'une entité unique

a.1) Localisations absolues

a.1.1) Localisations absolues à l'aide d'une préposition spatiale simple

Nous proposons l'algorithme suivant pour calculer les références effectuées à l'aide de

la préposition “de droite” :

<p>A1 : Localisation absolue d’une entité unique à l’aide d’une préposition simple. Exemple : “détruis le cercle de droite”.</p>
<p>tant que (aucun objet du type recherché ne rencontre l’aire de recherche) (figure 3.23.(a)) agrandir cette aire à partir de la droite (figure 3.23.(b)).</p>

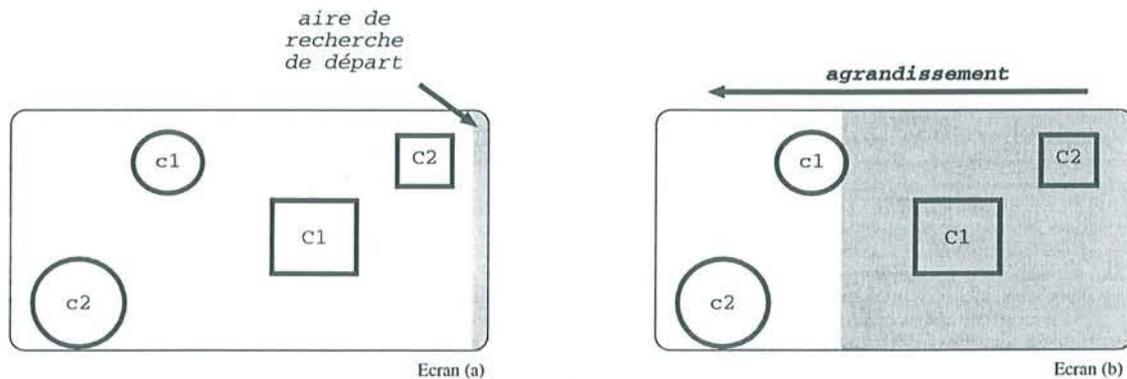


FIG. 3.23 – Modèle de calcul des référents “absolu” pour la préposition “de droite”.

Remarques :

1. Cet algorithme s’applique à tout l’écran. Nous verrons par la suite qu’il est parfois nécessaire de ne l’appliquer qu’à une sous-partie de ce dernier.
2. Lors de la présentation de notre contribution au paragraphe 3.5.2, nous avons pu démontrer que l’on pouvait astucieusement unifier le cas de la référence définie avec celui de la référence indéfinie. En effet, une requête du type “détruis l’un des carrés de droite” est satisfaite quel que soit le carré détruit, pourvu que le carré détruit se situe effectivement à droite. L’algorithme A1 traite alors le cas de la référence indéfinie comme une référence définie et renvoie en fait le carré qui se situe le plus à droite. Un problème demeure cependant si l’aire de recherche isole simultanément deux entités en même temps, voir la figure 3.24 pour la requête “détruis l’un des carrés de droite”. L’aire de recherche isolera en effet simultanément les carrés C6 et C7. La solution que nous préconisons en pareille situation consiste à renvoyer l’une des cibles trouvées au hasard. On pourra remarquer que cette ambiguïté entre deux ou plusieurs cibles n’apparaît pas pour le problème de la référence définie en vertu du principe de pertinence. En effet, si l’utilisateur énonce par exemple : “détruis le carré de droite”, il n’y a nécessairement, pour lui, pas d’ambiguïté quant à la cible concernée ce qui laisse présager que la recherche à partir des aires de recherche se fera sans ambiguïté.

a.1.2) Localisations absolues à l’aide d’une préposition spatiale composée

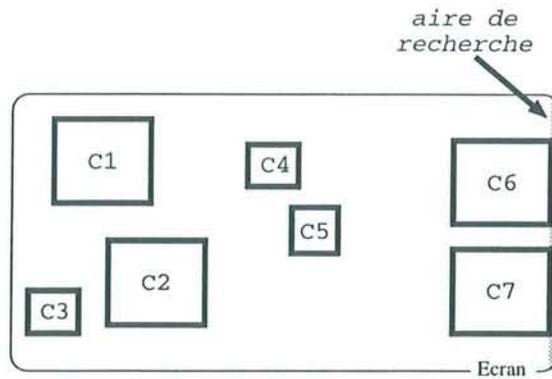


FIG. 3.24 – Ambiguïté possible pour les références indéfinies.

Pour la préposition “en haut à droite”, la recherche est effectuée à l’aide de la procédure suivante :

A2: Localisation absolue d’une entité unique à l’aide d’une préposition composée.
 Exemple : “détruis le carré qui se trouve en haut à droite”.
tant que (aucun objet du type recherché ne rencontre l’aire de recherche)
 (figure 3.25.(a))
 agrandir l’aire de recherche en partant en haut à droite, en utilisant des droites perpendiculaires à la diagonale de l’écran (figure 3.25.(b)).

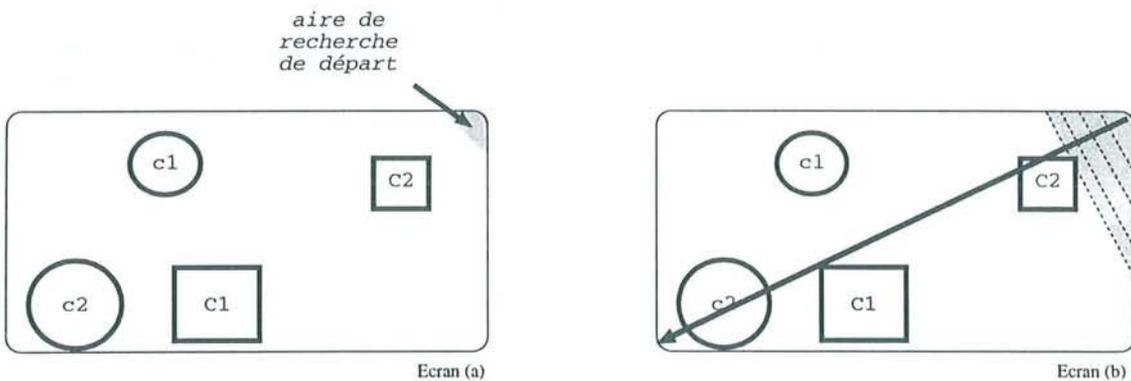


FIG. 3.25 – Modèle de calcul des référents “absolu” pour la préposition “en haut à droite”.

a.2) Localisations relatives

a.2.1) Localisations relatives par rapport à une entité unique

a.2.1.1) Localisations à l’aide d’une préposition spatiale simple

Comme mentionné précédemment, il nous faut exhiber un objet unique. Cela est réalisé pour la préposition “à droite de” à l’aide de l’algorithme suivant :

A3 : Localisation relative d’une entité unique par rapport à un site unique à l’aide d’une préposition simple. Exemple: “détruis le cercle qui se trouve à droite du carré”.

tant que (*aucun objet du type recherché ne rencontre l’aire de recherche*) et
 ($\text{angle}(\text{horizontale}, L1) < 45^\circ$) et ($\text{angle}(L2, \text{horizontale}) < 45^\circ$)
 agrandir progressivement l’aire de recherche (figure 3.26.(a)).

si (*aucun objet ne fait partie de l’aire de recherche*) **alors**
 agrandir progressivement l’aire de recherche par la droite (figure 3.26.(b)).

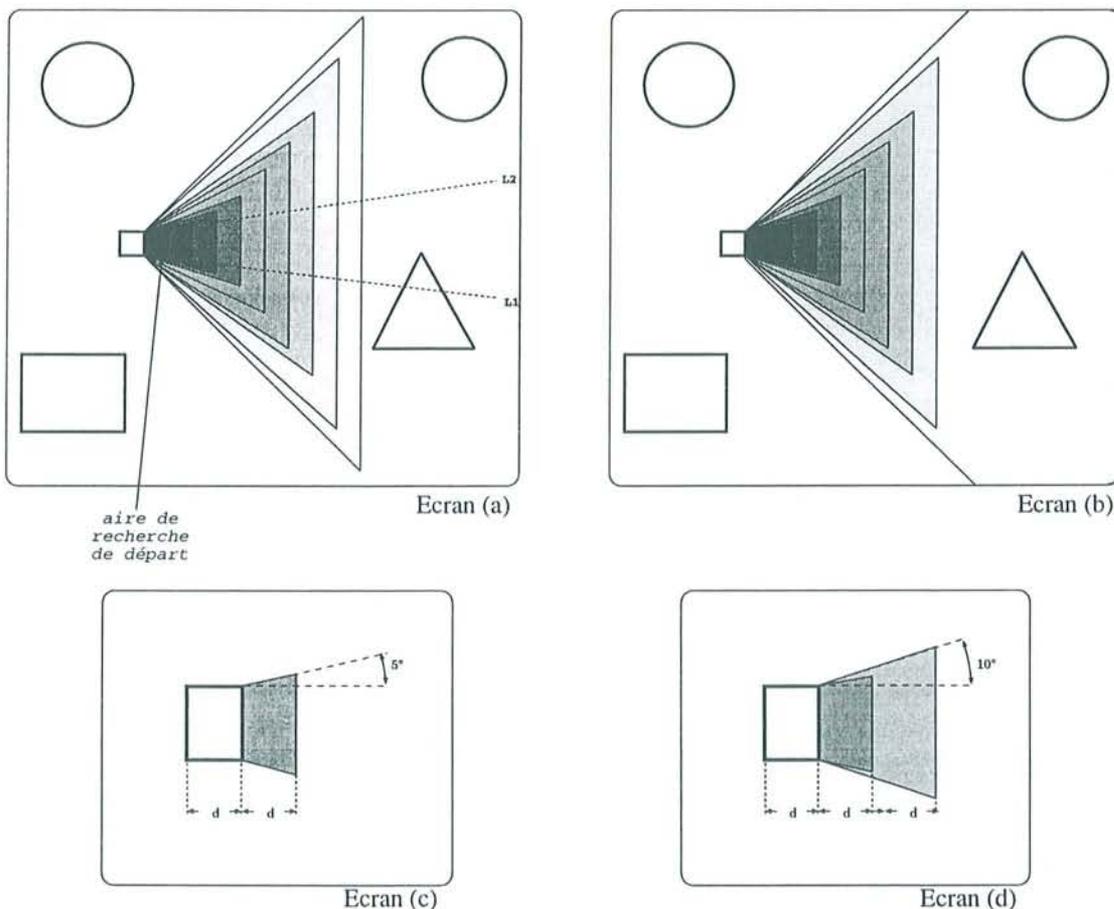


FIG. 3.26 – Modèle de calcul des référents “relatif” pour la préposition “à droite de”.

- Détails de construction des aires de recherche :

- On élabore tout d’abord le rectangle englobant le site qui est facile à construire quel que soit l’objet graphique considéré (triangle, cercle, carré et rectangle *a fortiori*).

- On agrandit alors à chaque fois l'aire de recherche horizontalement et verticalement. L'extension horizontale se fait à chaque fois d'une distance d (cette distance étant le minimum entre la hauteur et la largeur du rectangle englobant le site, figure 3.26.(c)). L'extension verticale est réalisée à l'aide de l'angle entre L1 et l'horizontale et l'angle entre l'horizontale et L2 qui seront augmentés à chaque fois de 5° (figures 3.26.(c) et 3.26.(d)) jusqu'à atteindre pour chacun d'eux 45° (figure 3.26.(b)).
- Dans une dernière étape, on étend progressivement l'aire de recherche vers la droite à chaque fois de la distance d (figure 3.26.(b)).

• Remarques :

1. Certaines valeurs d'agrandissement des aires de recherches présentent un caractère *ad hoc*, en fait nous les avons déterminées à partir de l'analyse d'un petit corpus d'écrans représentant des objets graphiques que nous nous sommes constitué.
2. L'angle de départ entre L1 et l'horizontale ainsi qu'entre L2 et l'horizontale est au maximum de 45° . Cet angle n'est, de notre point de vue, pas empirique mais correspond au contraire à la séparation qui peut être faite entre la zone d'acceptabilité de la préposition "à droite de" des zones d'acceptabilité définies par les prépositions "au-dessus de" et "en dessous de".

a.2.1.1) Localisations à l'aide d'une préposition spatiale composée

Voici l'algorithme qui effectue la recherche de l'entité spécifiée pour la préposition "en haut à droite de" :

A4 : Localisation relative d'une entité unique par rapport à un site unique à l'aide d'une préposition composée. Exemple : "détruis le cercle qui se trouve en haut à droite du rectangle".

tant que (*aucun objet du type recherché ne rencontre l'aire de recherche*) et ($\text{angle}(\text{horizontale}, L1) < 90^\circ$) et ($\text{angle}(\text{verticale}, L2) < 90^\circ$)
agrandir progressivement l'aire de recherche (figure 3.27.(a)).

si (*aucun objet ne fait partie de l'aire de recherche*) **alors**
agrandir progressivement l'aire de recherche de la distance d ,
verticalement et horizontalement (figure 3.27.(b)).

• Détails de la construction des aires de recherche :

- La distance d est à nouveau définie comme étant le minimum entre la largeur et la longueur du rectangle englobant.
- L'angle de départ entre L1 et L2 est de 30° . Cet angle est empirique et repose sur le corpus que nous nous sommes constitué.

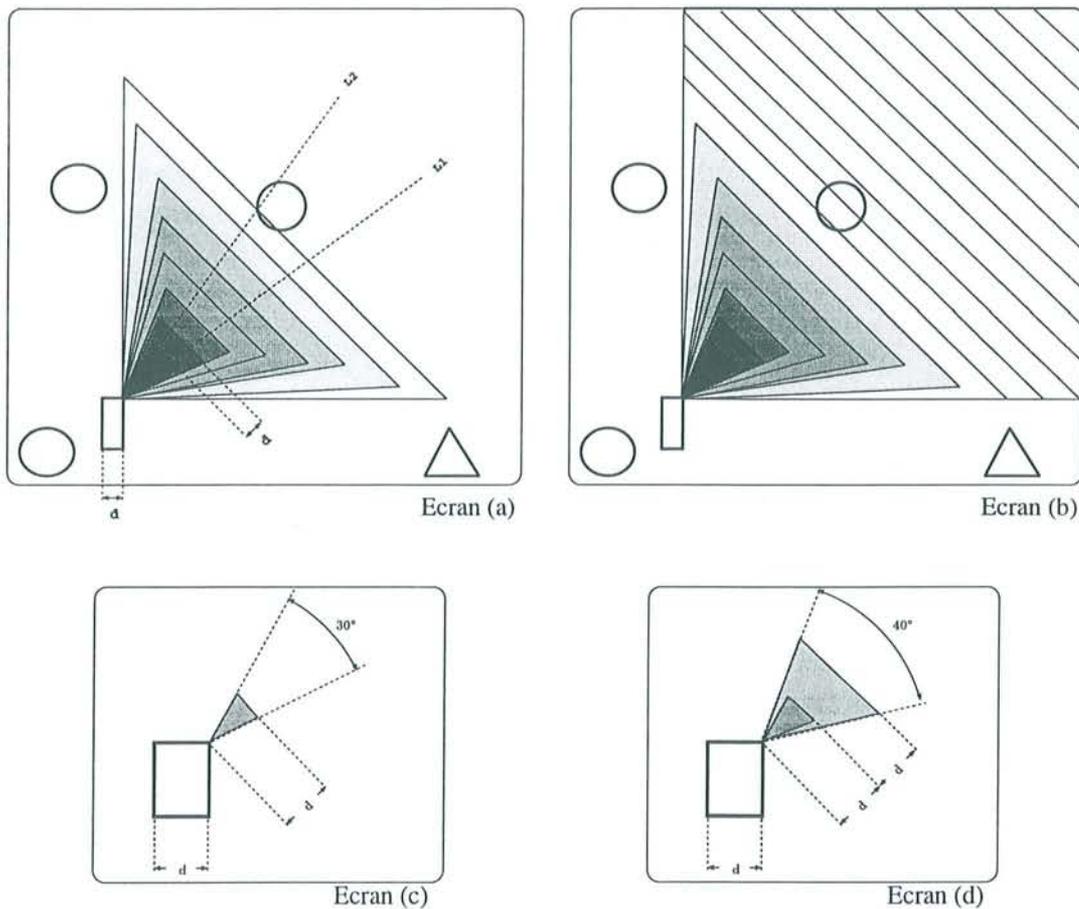


FIG. 3.27 – *Modèle de calcul des référents “relatif” pour la préposition “en haut à droite de”.*

- L’angle entre L1 et L2 est agrandi à chaque pas de 10° ainsi que la portée du nuage de localisation en termes de distance par rapport au site de la distance d (figures 3.27.(c) et 3.27.(d)).

a.2.2) Localisations relatives par rapport à un groupe d’entités

Pour ce faire, nous proposons d’appliquer les deux classes d’algorithmes que nous venons d’exposer ci-dessus (**A3** et **A4**) en prenant comme entité, jouant le rôle de site, le rectangle englobant les entités localisatrices. La détermination des entités localisatrices est supposée se faire sans ambiguïté comme nous l’avons supposé au paragraphe 3.5.2. Par ailleurs, la construction du carré englobant est facile car il suffit de prendre l’enveloppe convexe rectangulaire entourant les objets considérés, voir à ce titre, l’exemple de la figure 3.28 pour une requête du type : “*détruis le carré qui se trouve à droite des cercles*”.

Concernant cette dernière requête, on pourrait sans doute nous rétorquer qu’au sein d’un système multimodal, l’usager préférerait sans doute dire : “*détruis ce carré*” (+ geste de désignation) plutôt que “*détruis le carré qui se trouve à droite des cercles*”. Cependant, comme nous l’avons mentionné précédemment, notre souci est de laisser le maximum de liberté à l’utilisateur en lui laissant la possibilité d’utiliser de telles requêtes.

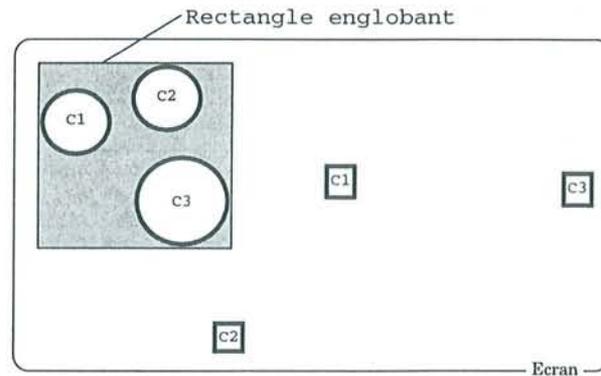


FIG. 3.28 – Modèle de calcul des référents pour une localisation à partir d'un groupe d'entités.

b) Localisations d'un groupe d'entités

Alors que précédemment il nous fallait retrouver une entité unique, à présent, la localisation porte sur un groupe d'objets. Comme nous allons le voir tout de suite, celui-ci possède nécessairement une homogénéité perceptive, en particulier, les entités recherchées se détachent clairement des autres par le fait qu'elles sont proches les unes des autres (voir par exemple la figure 3.29.(a) qui représente une succession de carrés sur laquelle on dirait par exemple “détruis les carrés de droite”).

Le fait de parler de groupes homogènes perceptifs rejoint la structuration spontanée de l'univers perceptif telle qu'elle est apparue au sein de la Gestalt. Delay et Pichot [1984] reprennent ces travaux et parlent de la tendance spontanée que les humains ont à structurer les éléments perceptifs isolés en *formes* ou *Gestalten*³². Ainsi, si nous regardons une ligne de points inégalement espacés, nous avons tendance à percevoir des groupes de points. La figure 3.30 illustre ce propos étant donné que les 32 points tendent spontanément, du fait de leur disposition spatiale, à être perçus comme quatre groupes de huit points chacun.

Briffault [1992] a proposé dans sa thèse un algorithme qui permet de construire les groupes homogènes perceptifs tels qu'ils devraient être mis en évidence à la figure 3.30. Son principe de fonctionnement est le suivant : chaque objet est relié à l'objet dont il est le plus proche par un arc Pp (plus proche). Un arc existe alors entre un objet A et un objet B si et seulement si on a soit $A Pp B$ ou $B Pp A$. Ainsi, dans l'exemple de la figure 3.29.(a), l'objet 1 est plus proche de l'objet 3 que de n'importe quel autre objet. De même l'objet 6 est plus proche de l'objet 5 que de n'importe quel autre. Ces arcs permettent alors de structurer la scène affichée sous forme d'un ensemble de sous-graphes. Deux objets A et B sont placés dans un même sous-graphe s'il existe un chemin formé d'un ou plusieurs arcs qui permet d'aller de l'un à l'autre (voir figure 3.29.(b), ces sous-graphes seront dénommés dans la suite *groupes de proximité*). Par ailleurs, si l'on demande à des sujets un regroupement plus précis, on obtient les groupements de la figure 3.29.(c). Ces groupements sont obtenus en plaçant dans une même catégorie les objets A et B qui

³². Nous reviendrons plus en détail sur ces phénomènes de structuration du champ visuel au prochain chapitre.

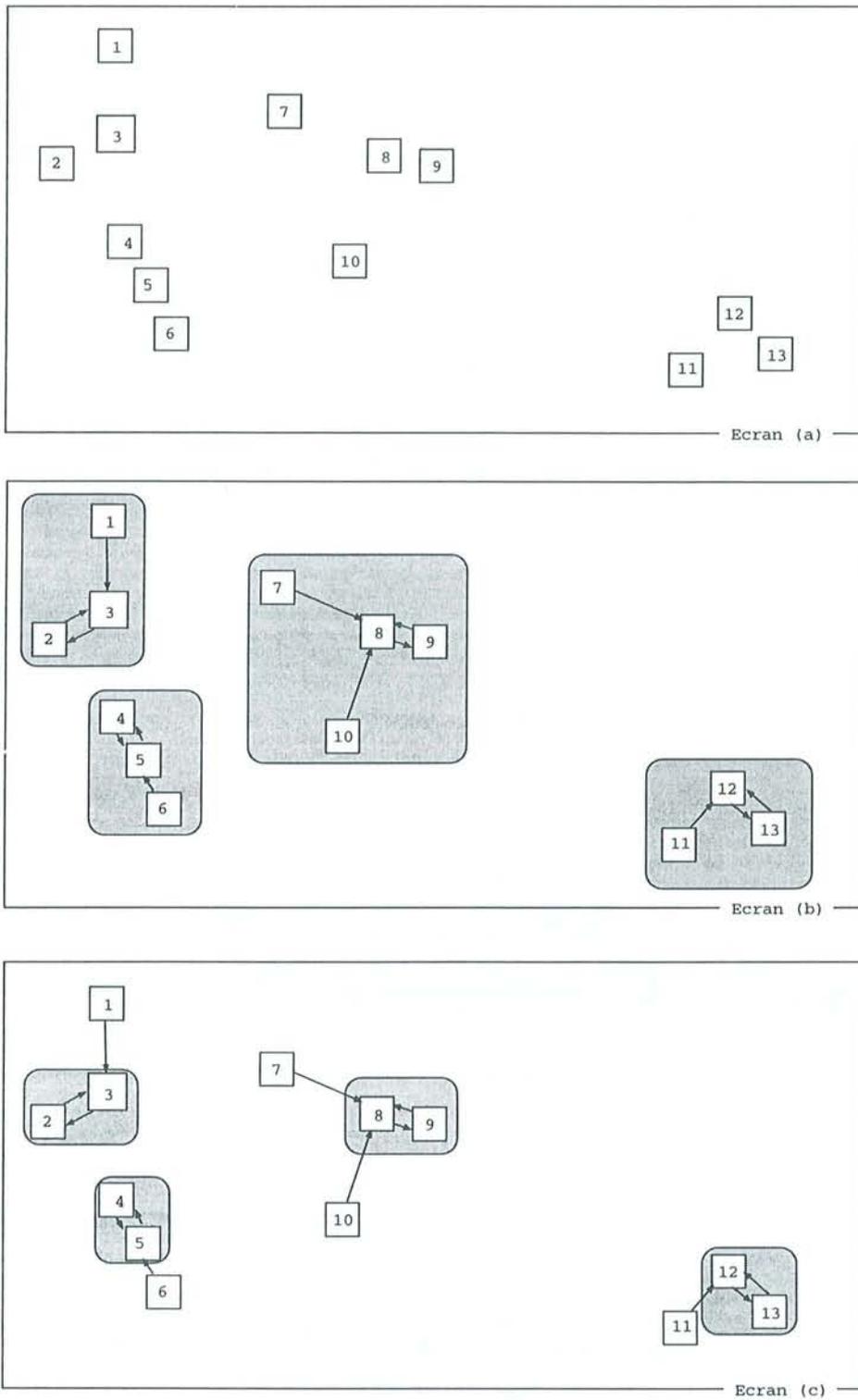


FIG. 3.29 – Construction des groupes homogènes perceptifs selon Briffault.

vérifient : $A Pp B$ et $B Pp A$.

Cet algorithme présente quelques limitations. En effet, prenons l'exemple de la requête



FIG. 3.30 – *Structuration spontanée d'un ensemble d'éléments (d'après [Delay et Pichot, 1984]).*

“détruis les carrés de droite” qui serait appliquée à la figure 3.31.(a). Le premier type de regroupement que propose Briffault aboutirait à un ensemble perceptif unique (figure 3.31.(b)) nous empêchant d'isoler un groupe homogène perceptif unique. Enfin, la solution qui consisterait à recourir à un regroupement perceptif plus précis ne fonctionne pas non plus car on isolerait alors le carré C4 des carrés C2 et C3, ce qui n'est pas souhaitable (voir les groupes hachurés à la figure 3.31.(b)).

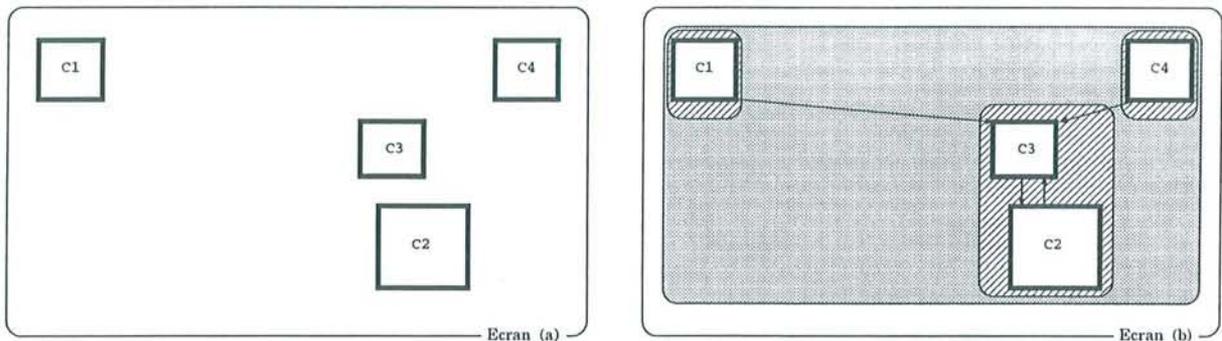


FIG. 3.31 – *Comment construire les groupes homogènes perceptifs?*

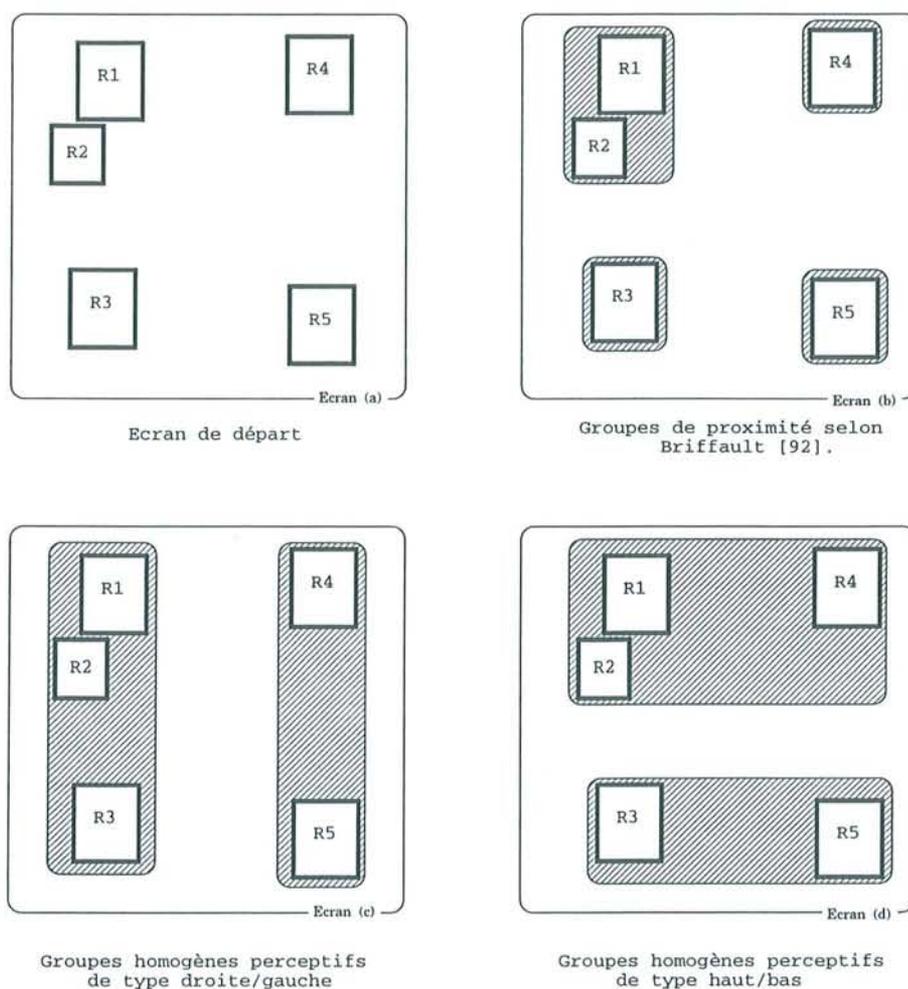
La figure 3.30 a mis en évidence le cas d'une structuration spontanée de l'environnement en groupes homogènes perceptifs, ces derniers nous apparaissent consciemment sans que nous ayons à faire le moindre effort. Ce type de structuration ne nous suffira cependant pas dans la suite. En effet, prenons le cas de la figure 3.32.(a) pour laquelle Delay et Pichot distinguaient quatre groupes perceptifs spontanés (voir figure 3.32.(b)). À présent imaginons que nous souhaitions faire la requête (107) ou la requête (108) :

Détruis les rectangles de droite. (107)

Détruis les rectangles qui se trouvent en bas. (108)

Il nous semble alors que la préposition “de droite” ou “en bas” induit une forte structuration de la scène. Ainsi, dans le premier cas, R1, R2 et R3 sont opposés à R4 et R5 tandis que dans le second R1, R2 et R4 s'opposent à R3 et R5. Cet exemple démontre qu'il y a une très forte interaction entre la langue et la perception. L'utilisateur projetant pour ainsi dire un point de vue sur la scène d'objets affichée ce qui lui permet de structurer cette dernière.

Comme nous allons l'illustrer ci-après, pour prendre en compte ce type de contraintes,

FIG. 3.32 – *Groupes de perception homogènes.*

nous proposons de recourir respectivement à des groupes homogènes perceptifs de type *droite/gauche* (figure 3.32.(c)) et de type *haut/bas* (figure 3.32.(d)). La construction de ces groupes se fait en adoptant les algorithmes exposés à l’annexe C. Nous allons décrire à présent comment interviennent ces groupes homogènes perceptifs pour le problème de la localisation absolue d’un groupe d’entités.

b.1) Localisations absolues d’un groupe d’entités

Qu’il s’agisse d’une localisation absolue à l’aide d’une préposition simple (comme “à droite de” par exemple) ou d’une localisation absolue effectuée à l’aide d’une préposition spatiale composée, il suffit d’appliquer les algorithmes **A1** et **A2** en cherchant à isoler un groupe homogène perceptif unique.

Comme mentionné plus haut et à l’annexe C, la construction des groupes homogènes perceptifs diffère suivant la préposition spatiale qui est mise en jeu :

- Pour les localisations à l’aide des prépositions “de droite” et “de gauche” on utilise

la construction des groupes perceptifs de type *droite/gauche* tels qu'ils apparaissent à la figure 3.32.(c) et dont l'algorithme d'obtention est décrit à l'annexe C.

- Pour les localisations à l'aide des prépositions "*en haut*" et "*en bas*", il faut utiliser les groupes perceptifs de type *haut/bas* tels qu'ils apparaissent à la figure 3.32.(d), voir l'annexe C pour l'algorithme de construction de ces groupes.
- Enfin, pour les localisations restantes, nous proposons de recourir à l'élaboration des groupes perceptifs suivant la méthode qui a été exposée pour la figure 3.32.(b) en y adjoignant la notion de distance seuil de regroupement (voir Annexe C). Leur construction se fait sur la base des groupes homogènes perceptifs de Briffault tels qu'ils ont été mis en évidence à la figure 3.29.(b) où deux objets A et B font partie de même groupe homogène perceptif si $A Pp B$ et $B Pp A$.

b.2) Localisations relatives d'un groupe d'entités

Il suffit simplement d'appliquer les algorithmes **A3** et **A4** en construisant les groupes homogènes perceptifs adéquats. Ainsi, si nous nous trouvons dans le cas d'une localisation par rapport à un groupe d'objets, il nous suffit de construire le rectangle englobant et d'appliquer les algorithmes **A3** et **A4** par rapport à ce dernier.

c) Bilan de ce modèle

Parmi les avantages de notre approche, nous noterons que le modèle développé ci-dessus est minimal au sens où il se fonde sur le principe d'économie cognitive. En effet, il n'est pas capable de donner précisément la position d'un objet A par rapport à un objet B. Il est simplement capable de retrouver l'(les) objet(s) qui a(ont) été désigné(s) par l'utilisateur³³. Notre modèle se fonde sur une approche qualitative, en particulier, l'(es) entité(s) spécifiée(s) est(sont) recherchée(s) par discrétisation de l'espace. Nous renvoyons le lecteur au chapitre 1, paragraphe 1.4.3 où nous défendons les vertus des approches qualitatives par rapport aux approches quantitatives.

D'autre part, il est important de noter que le modèle que nous proposons présente l'avantage d'apporter une solution au problème de la référence absolue, de la référence plurielle et de la portée localisatrice limitée du site. En effet, pour le problème de la référence absolue, pour laquelle l'ensemble des modèles décrits n'étaient pas prévus³⁴, notre approche, qui procède par extension progressive permet d'isoler correctement l'(les) entité(s) recherchée(s). Il suffit de prendre la figure 3.33.(a) pour la requête "*détruis le carré de droite*" où notre approche réussit à isoler correctement le carré C2 en procédant par extension de la zone de recherche vers la gauche jusqu'à isoler un carré unique (figure 3.33.(b)).

En ce qui concerne la référence plurielle, nous avons pu démontrer l'intérêt d'introduire la notion de groupe homogène perceptif. Là encore, notre modèle peut apporter certaines

33. Cette remarque rejoint la distinction que nous faisons plus haut entre un énoncé descriptif et un énoncé purement référentiel.

34. Excepté le modèle de Wazinski, cependant, nous avons pu démontrer que ce modèle présentait tout de même certaines limites pour ce type de localisation.

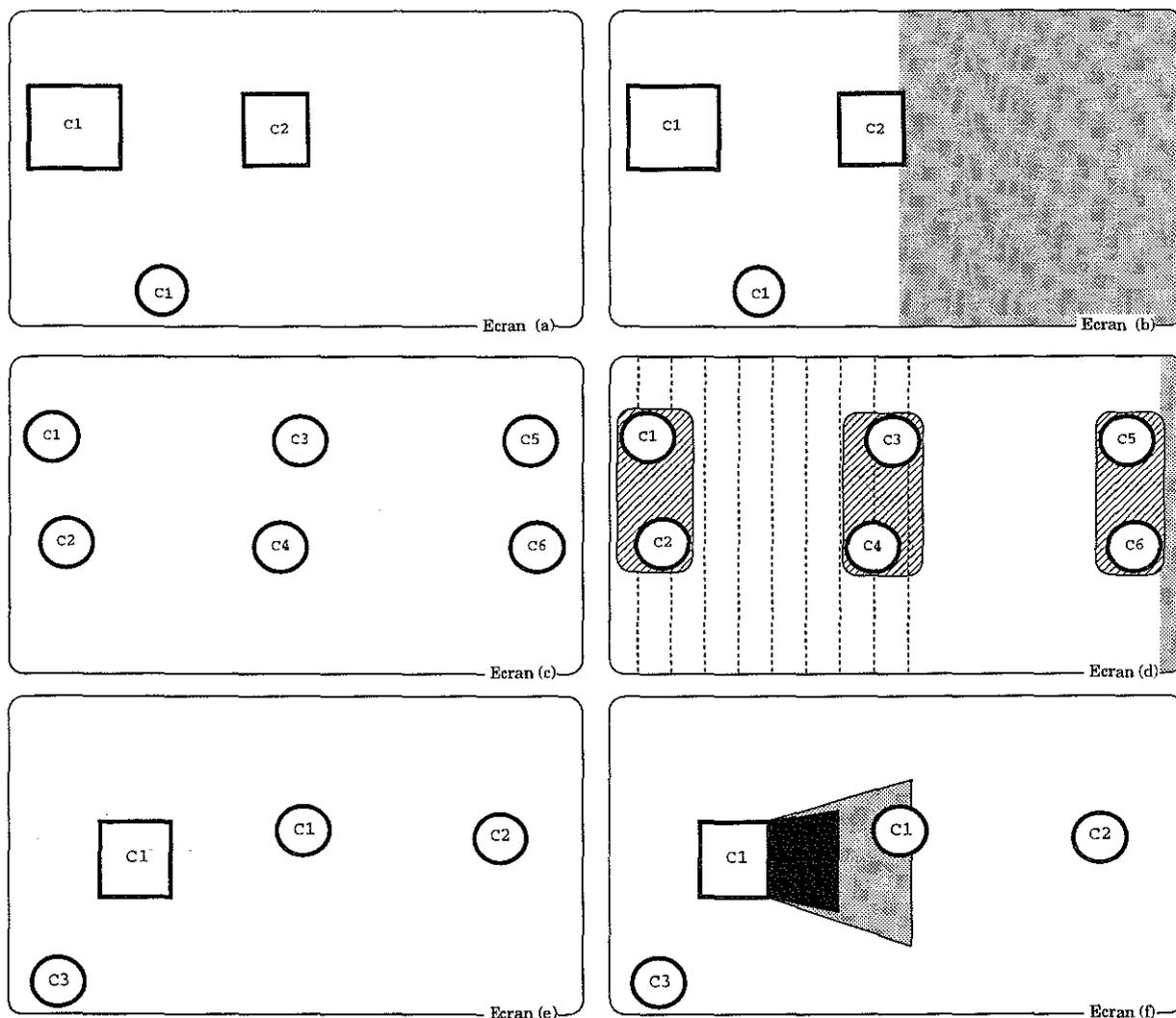


FIG. 3.33 – Apports du modèle développé.

solutions. Pour ce faire, il suffit de considérer la figure 3.33.(c) pour la requête “*détruis les cercles de droite*” où une recherche en partant de la droite, jusqu’à atteindre un groupe homogène perceptif unique de type *droite/gauche*, nous conduira à isoler les cercles C5 et C6 (figure 3.33.(d)).

Finalement, pour le problème spécifique de la portée localisatrice limitée du site, l’extension progressive permet là encore d’apporter certaines solutions. En particulier, à la figure 3.33.(e) où il s’agit d’isoler le cercle C1 par une requête du type “*détruis le cercle qui se trouve à droite du carré*”, l’extension de la zone de localisation nous conduira à retrouver effectivement le cercle voulu (figure 3.33.(f)).

Parmi les limites de notre approche, on se rappellera que nous nous étions sciemment placés dans le cas où les objets graphiques ne se chevauchaient pas. En particulier les requêtes mettant en jeu la préposition “*devant*” ou “*derrière*” ne sont pas traitées.

Enfin, les algorithmes que nous avons exposés ne peuvent gérer une requête du type

“détruis le cercle qui se trouve à droite des carrés” qui serait effectuée à la figure 3.34 par exemple. En effet, nos algorithmes y isoleraient le mauvais groupe localisateur (qui serait constitué de l'ensemble des carrés C1, C2, C3 et C4) ce qui nous empêcherait de retrouver le cercle C1. Il serait donc nécessaire ici de pouvoir distinguer les différents groupes homogènes perceptifs pouvant jouer le rôle de site en cas de référence plurielle par rapport à plusieurs sites.

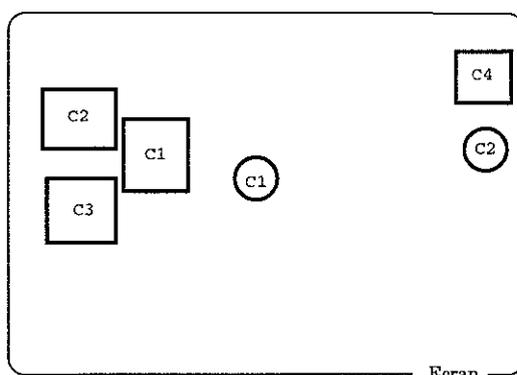


FIG. 3.34 – Ambiguïté dans la détermination du groupe localisateur.

On pourra remarquer que cet exemple s'inscrit dans la problématique plus générale de l'éventuelle ambiguïté du site que nous avons délibérément écartée de nos préoccupations dans ce chapitre.

3.8 De la nécessité d'introduire la notion de cadre

Par défaut, la recherche du(des) entités référencée(s) se fait sur la totalité de l'écran au départ. Cependant, bien souvent, elle pourrait (et devrait) se faire sur une partie de ce dernier. Il suffit de prendre l'exemple de la figure 3.35 sur laquelle on effectue les requêtes suivantes :

Agrandis les carrés de gauche. (109)

Détruis celui de droite. (110)

Un système classique proposerait de détruire le carré qui se trouve le plus à droite, en l'occurrence, le carré C4. De notre point de vue, cela n'est pas si sûr, le locuteur a très bien pu vouloir désigner le carré C1 qui vient juste d'être agrandi (il existe en fait une ambiguïté entre les deux carrés C1 et C4). Ce dernier exemple est à rapprocher de celui que nous avons proposé au paragraphe 3.4.3 :

Déplace les fenêtres vertes. (111)

Mets la fenêtre la plus à gauche en arrière plan. (112)

Le fait de pouvoir calculer, à la phrase (110), le bon référent spatial nous imposera un suivi discursif. Nous nous intéresserons à ce type de préoccupations au chapitre suivant.

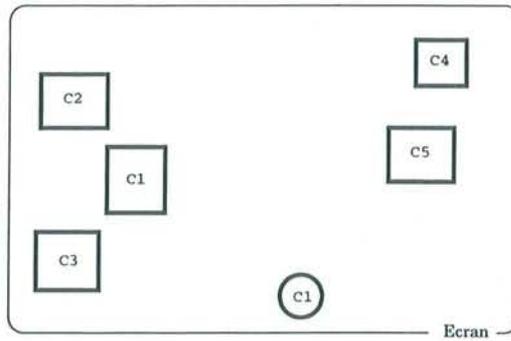


FIG. 3.35 – De la nécessité des cadres.

3.9 Conclusion

Ce chapitre illustre ce qu'avait pressenti Klein [1993] et bien d'autres auteurs avant elle, à savoir que le problème de la référence est un processus cognitif complexe qui met en jeu l'interaction d'une multitude de facteurs.

L'étude spécifique de la référence spatiale portant sur un système d'affichage graphique nous a conduit à plusieurs réflexions et conclusions. Notre première réflexion a porté sur la nécessité, dans le cadre de la réalisation de systèmes de dialogue multimodaux, d'améliorer la composante spatiale purement langagière de tels systèmes. En effet, si l'importance et l'engouement pour le développement de systèmes de dialogue homme-machine multimodal est de plus en plus manifeste ([Carré *et al.*, 1991], [Bourguet, 1992], [Poirier et Lefebvre, 1993] et [Gaildrat *et al.*, 1993] pour ne citer qu'eux), nous pensons que les systèmes actuels gagneraient en efficacité à améliorer les modèles d'interprétation de chacune de ces modalités, même indépendamment les unes des autres [Schang, 1994]. Ce chapitre vise donc à autoriser certains types de références spontanées de l'utilisateur, en l'occurrence certaines références spatiales élaborées. Notre contribution va donc dans le sens de la réalisation d'une interface moins contraignante pour l'utilisateur.

L'étude des difficultés rencontrées par les modèles actuels pour prendre en compte les références spatiales nous a conduit à définir notre propre modèle. Ce dernier exploite au maximum les différences que nous avons pu mettre en évidence entre un système qui chercherait à faire une description d'une scène d'objets affichée à l'écran par opposition à un système qui chercherait simplement à faire référence à certains de ces objets.

Après une étude critique des apports possibles de la notion de saillance et de prototype, nous avons proposé un modèle qui traite les expressions spatiales composées de certaines prépositions spatiales projectives comme "à gauche de", "au-dessus de", "en dessous de", "de gauche", etc. Ce modèle apporte des éléments de réponse pour le problème de la référence absolue, celui de la référence plurielle et enfin celui de la portée localisatrice limitée du site de par son principe de fonctionnement à base d'extension progressive de l'aire de recherche. Nous rappelons ici le principe de fonctionnement des algorithmes que nous avons proposés au paragraphe 3.7. Chaque algorithme étend progressivement sa zone de recherche à partir du(des) site(s) jusqu'à isoler une cible ou un groupe de

cibles présentant la particularité d'être homogène perceptivement. Il semble par ailleurs que ce mode d'extension progressif de la zone de recherche soit extensible à d'autres prépositions telles que "*près de*", "*derrière*" en changeant simplement l'extension de l'aire de recherche. En revanche, le modèle proposé n'est pas capable de suivre le contexte au cours du discours ce qui nous aurait permis d'appliquer nos schémas sur une sous-partie de l'écran lorsque cela s'avérait nécessaire.

Par ailleurs, au sein des algorithmes, certaines données sont empiriques comme par exemple l'angle de 30° qui définit l'aire de recherche de la préposition "*en haut à droite de*". Ces choix ont été déterminés à partir de l'analyse d'un petit corpus d'écrans représentant des objets graphiques que nous nous sommes constitué.

Nous avons également insisté dans ce chapitre sur le caractère localisateur de la cible qui a par trop longtemps été occulté, cette dernière ayant également un rôle à jouer dans sa propre localisation.

Avec le recul, il nous paraît très difficile d'élargir la méthodologie que nous avons proposée pour l'étendre au problème général de la modélisation de l'espace en langage naturel. En effet, comme nous avons pu le voir dans ce chapitre, faute d'une réflexion préalable suffisamment générale sur l'espace, nous avons été amenés à apporter rapidement de nombreuses restrictions à notre modèle ce qui le place dans une perspective d'usage très spécialisée. Ce dernier point justifie donc l'étude plus générale que nous allons entreprendre au chapitre suivant concernant les nombreux problèmes attenants à la modélisation de l'espace en langage naturel.

Chapitre 4

Vers un nouveau concept, celui de cadre

Sommaire

4.1	Introduction	126
4.2	De la nécessité des cadres	127
4.2.1	Une nécessité au niveau spatial	127
4.2.2	Une nécessité au niveau informatique	128
4.2.3	Conclusion	130
4.3	Définition et caractérisation de la notion de cadre	131
4.3.1	Définition générale	131
4.3.2	Les cadres et le problème de la référence spatiale	132
4.3.3	Tout est “ <i>alternative</i> ”!	133
4.3.4	Caractérisation des cadres de référence	135
4.3.5	Évolution discursive des cadres de référence	138
4.4	Les cadres de référence et l’existant	140
4.4.1	Les cadres de référence et l’approche de Vandeloise	140
4.4.2	Les cadres de référence et l’approche de Asher et Sablayrolles	142
4.4.3	Les cadres de référence et les aires de recherche de Pribbenow	143
4.4.4	Les cadres de référence et les univers spatiaux de Charolles	144
4.4.5	Conclusion	146
4.5	Connaissances nécessaires à l’interprétation d’une référence spatiale	147
4.5.1	Le nouveau cadre de référence	147
4.5.2	L’historique du dialogue et le lexique	159
4.5.3	Analyse syntaxico-sémantique	162
4.5.4	La perception	173
4.5.5	L’application	178

4.6	Principe de fonctionnement de la notion de cadre de référence	179
4.7	Application de la notion de cadre de référence à un système de représentation graphique	183
4.7.1	Introduction	183
4.7.2	Exemple détaillé	183
4.7.3	Le modèle de la notion de cadre de référence: un bilan	191
4.8	Prémises pour un modèle de traitement des énoncés de positionnement	192
4.8.1	Introduction	192
4.8.2	Vers un positionnement plus “cognitif”, le cas de “ <i>ici</i> ”	193
4.8.3	Quelques pistes pour formaliser les énoncés de positionnement	195
4.8.4	Conclusion	197
4.9	Autres validations des cadres de référence	198
4.9.1	Modèles informatiques, modèles de la cognition humaine: un cas de divorce?	198
4.9.2	Validation psychologique des cadres de référence	199
4.9.3	Cadres de référence et ambiguïté	200
4.9.4	Les cadres de référence et la transitivité	202
4.9.5	Liens avec le temps	202
4.10	Conclusion	203

4.1 Introduction

Au vu du caractère très spécifique des résultats obtenus lors du précédent chapitre, il nous a semblé nécessaire de mener une réflexion plus générale sur la modélisation de l'espace en langage naturel.

Cette réflexion démontrera tout d'abord la nécessité d'introduire un nouveau concept, très général, qui est celui de *cadre*. Après avoir défini et précisé cette notion, nous l'instancierons pour le problème spécifique de la référence spatiale, ce qui nous amènera à considérer la notion de *cadre de référence*. Là encore, après avoir défini cette notion, nous en préciserons les fondements qui se déclinent suivant des spécificités perceptuelles, fonctionnelles et sémantiques. Nous insisterons sur la notion d'alternative qui semble relever, tout comme la notion de principe d'économie cognitive, de l'une des facettes de la cognition humaine.

Après avoir défini la notion de cadre de référence et mentionné ses différentes évolutions discursives possibles, nous chercherons à valider ce nouveau concept en quatre étapes. Nous essaierons dans un premier temps de le situer par rapport aux autres approches et modèles actuels. Dans un second temps, nous étudierons précisément de quelle manière les cadres de référence peuvent intervenir pour réaliser l'interprétation d'une référence spatiale au sein d'un système de dialogue homme-machine à forte composante graphique. Cette

partie nous fournira l'occasion de présenter à partir de quelles connaissances s'effectue l'élaboration d'un cadre de référence. Dans un troisième temps, nous fournirons les bases de ce que pourrait être un modèle de traitement des énoncés de positionnement reposant sur la notion de cadre de référence. Finalement, dans une dernière étape, nous nous intéresserons à la validité psychologique et linguistique des cadres de référence ainsi que leurs liens éventuels avec certains aspects temporels. Nous mènerons enfin une réflexion portant sur l'éventuelle nécessité d'une adéquation entre le fonctionnement d'un modèle informatique en regard du fonctionnement de la cognition humaine.

4.2 De la nécessité des cadres

4.2.1 Une nécessité au niveau spatial

Imaginons, pour l'instant, la notion de cadre comme étant une portion d'espace ou la réunion de portions d'espace permettant de ne pas considérer l'espace dans sa totalité. Le fait que nous nous restreignons spontanément à une sous-partie de l'espace environnant lorsque nous lisons un texte comportant des informations de nature spatiale ou lorsque nous dialoguons avec un interlocuteur, en échangeant là encore des informations spatiales, paraît naturel. Pour illustrer ce propos, il suffit de prendre l'exemple du dialogue suivant qu'entretiendraient Pierre et Marie pendant le repas de midi :

Paul: *La viande, tu l'as achetée à la boucherie du centre?* (113)

Marie: *Oui, le supermarché n'était pas encore ouvert.* (114)

Paul: *Ah oui, elle est beaucoup plus tendre. Tiens, est-ce que c'est toi qui a fait cette tâche sur la nappe Marie?* (115)

Marie: *Je crois que ça vient de la bouteille de jus d'orange.* (116)

Paul: *En effet, regarde, il y a un trou sur le côté.* (117)

Au moment où Pierre énonce (113), Marie s'est pour ainsi dire "transportée" à la boucherie, elle s'est construit un cadre spatial qui englobe cette dernière. Ce cadre chemine ensuite au fil du dialogue. Enfin, au moment où Pierre énonce (117), il va de soi que Marie a complètement oublié le cadre spatial qui se situait autour de la boucherie (au moins temporairement), elle s'est complètement focalisée sur la bouteille de plastique et la nappe. *A fortiori* on peut imaginer qu'elle ne pense absolument pas à sa fille qui passe par exemple un examen à 200 kilomètres de là.

Ce petit exemple démontre la grande adaptabilité de l'esprit humain. En effet, que ce dernier se trouve impliqué dans un dialogue, ou qu'il se préoccupe de la compréhension d'un texte écrit, à un instant donné, il ne se focalisera que sur l'information pertinente, en évacuant toutes les informations inutiles. Cette dernière remarque fait écho à la notion de principe d'économie cognitive que nous défendons dans le premier chapitre.

Voyons à présent, dans le domaine de l'informatique, en quoi il peut être également nécessaire de se restreindre à une sous-partie de l'espace, en l'occurrence, une sous-partie de l'écran de l'ordinateur.

4.2.2 Une nécessité au niveau informatique

a) Les cadres et les expressions définies

Il semble qu'il faille relativiser la notion d'unicité qui est habituellement associée à l'interprétation des descriptions définies. En effet, cette unicité ne doit être comprise qu'à l'intérieur du contexte local qui est activé à un instant donné du discours ou du dialogue et non pas globalement au niveau de celui-ci. Ainsi, si nous reprenons le cas d'un utilisateur aménageant une pièce dans le cadre de l'expérience de type Magicien d'Oz de Mignot [1993]¹, l'énoncé "*déplace le canapé, il me gêne*" désigne un meuble unique, quand bien même il y aurait d'autres canapés visualisés à l'écran.

b) Les cadres et la portée localisatrice du site

Vandeloise, en se fondant sur la notion de cible/site à laquelle il adjoint celle de ressemblance de famille, nous propose les caractérisations suivantes des prépositions "*devant/derrière*" :

A est "*devant/derrière*" B si A se trouve du côté positif/négatif déterminé par l'orientation générale².

Cette définition n'est pas assez restrictive car elle ne limite pas assez le pouvoir localisateur du site. Prenons l'exemple de la phrase (118) où l'interlocuteur regarderait devant la voiture mais nécessairement à proximité de cette dernière.

*Oh! regarde, on dirait qu'il y a un billet de 500 Francs
là-bas, devant la Jaguar.* (118)

Plus généralement la portée d'une préposition dépend de l'objet sur lequel elle s'appuie, ainsi la portée de "*devant*" dans "*devant l'église*" couvre une zone plus vaste que dans "*devant la voiture*" par exemple. Ce fait est indirectement pris en compte dans le système Soccer [Schirra, 1993] par les nuages de probabilités qui s'appuient sur le bord des objets dépendant de leur forme et de leur étendue [Schirra et Stopp, 1993].

La portée limitée du site fait partie de l'une des contraintes que nous souhaiterions exprimer et traduire par la notion de cadre. Notons enfin que cette étendue peut être limitée davantage, en particulier par des entités contiguës. Prenons l'exemple d'un joueur

1. On trouvera à la page 204 de ce chapitre un extrait du dialogue spontané d'un utilisateur dans une telle situation de commande ainsi que le salon d'aménagement qui était visualisé sur l'écran de l'ordinateur. Le caractère agrammatical des phrases tient au fait que ces dernières proviennent de la transcription écrite des énoncés oraux et gestuels.

2. *Orientation générale* : ressemblance de famille dont les principaux traits sont : *la direction frontale, la direction du mouvement, la direction du regard, la direction des autres organes de perception : le goût, l'ouïe, etc.*

qui se trouve sur un terrain de football à proximité d'une balle qui est située dans le but (voir figure 4.1.(a)). On préférera sans aucun doute (119) à (120) :

La balle est dans le but. (119)

La balle est devant le joueur. (120)

Par défaut, la portée de “*devant le joueur*” est limitée par la discontinuité spatio-fonctionnelle du bord du terrain de football, limitation qui est renforcée par l'introduction d'un autre site (le but) qui limite la portée localisatrice du joueur de football (voir la zone hachurée à la figure 4.1.(b)). Il est important de noter que cette limitation est une limitation par défaut. En effet, on peut très bien imaginer un point de vue différent, qui ferait complètement disparaître les limites du terrain de football. À titre d'illustration, on peut considérer le cas de deux bureaux de travail contigus. Dans ce cas, les murs de ces bureaux constituent de fortes limitations spatio-fonctionnelles. Pour autant, si vous vous trouvez dans l'un d'eux avec un collègue, ce dernier peut très bien vous dire “*attention, ton chef est derrière toi*”, alors que votre supérieur se trouve dans l'autre bureau. Dans ce cas, le mur qui vous sépare de votre chef a pour ainsi dire disparu.

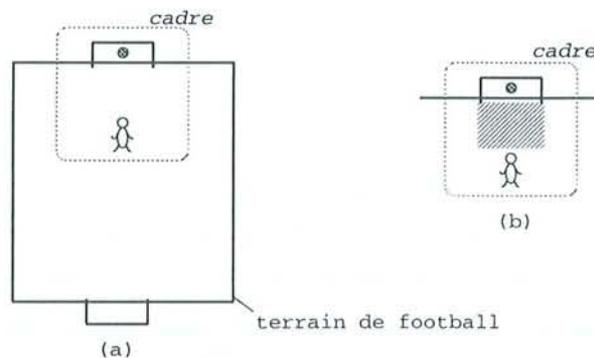


FIG. 4.1 – Limitations de la portée de la préposition “*devant*”.

Nous reviendrons sur ces problèmes de limitations de la portée localisatrice d'un site par un autre site qui serait placé à proximité du premier ; Pribbenow [1993] s'est en effet intéressée à ces différents problèmes comme nous le verrons dans la suite de ce chapitre.

c) Applications finalisées et cadres

La notion de cadre a un rôle important à jouer au sein des applications finalisées comme nous allons l'illustrer pour le cas d'un système de multifenêtrage. Ainsi, prenons l'exemple de la phrase “*mets un bouton nommé Print à droite du bouton Ok*” énoncée pour l'écran représenté à la figure 4.2. Si l'on peut trouver la position verticale du bouton à l'aide d'informations prototypiques (en utilisant un prototype de la préposition “*à droite de*”) et si la taille du bouton peut être déduite de celui existant, en revanche, trois positions horizontales pour le bouton sont au moins possibles. Parmi ces trois possibilités, il faut qu'il ne reste plus que la première. La position 2 peut être rejetée d'emblée pour

des raisons fonctionnelles liées à la tâche (i.e. un bouton doit être placé dans une fenêtre et pas directement sur l'écran). La position 3 semble contradictoire avec la signification de la phrase qui établit une relation directe entre les deux boutons dans le même contexte. Il s'agit là encore de l'une des contraintes que nous tenterons d'exprimer à l'aide de la notion de cadre.

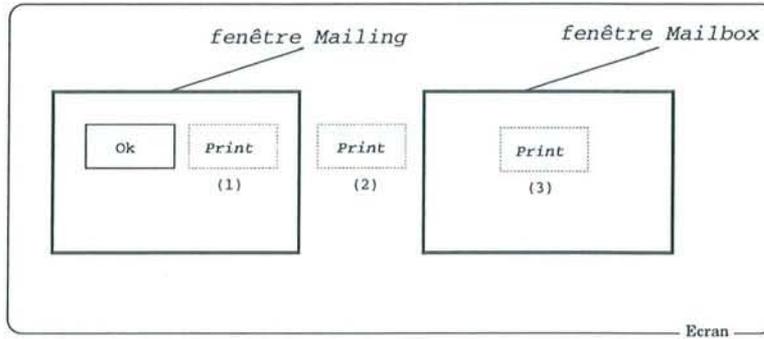


FIG. 4.2 – Nécessité des cadres dans un système de multifenêtrage.

Prenons à présent un second exemple, imaginons la suite de requêtes suivantes effectuées sur la figure 4.3 :

Mets le bouton back de la fenêtre Mailbox à gauche. (121)

Rapproche le bouton ok du bouton back. (122)

Un système informatique classique hésiterait entre les deux boutons “back”. Or, de notre point de vue, la référence au bouton “back” n’est pas ambiguë car à la phrase (122), il nous faut calculer en priorité les référents dans le cadre courant, qui évolue au fil du discours et se restreint à la fenêtre “Mailbox” après la requête de la phrase (121).

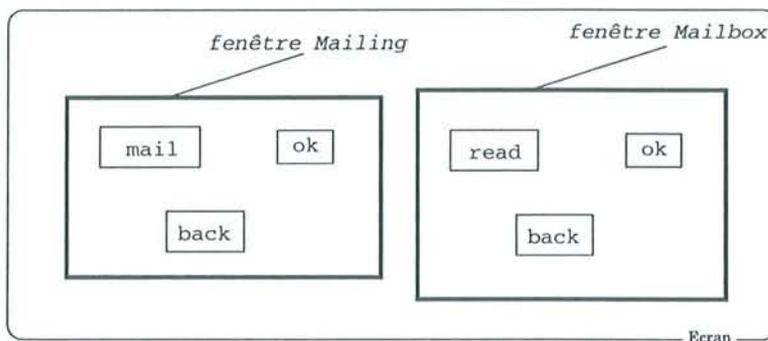


FIG. 4.3 – De la nécessité des cadres.

4.2.3 Conclusion

Au terme de ces différents paragraphes, nous pouvons commencer à esquisser certaines composantes sous-jacentes à la notion de cadre. Tout d’abord, le cadre aura

une étendue réduite, et ce pour trois raisons au moins. En premier lieu, son étendue dépendra fonctionnellement du site contenu dans le cadre. On a pu voir que le cadre englobant une église serait plus grand que celui englobant une voiture par exemple. En second lieu, la limitation de cette étendue pourra également provenir de discontinuités spatio-fonctionnelles comme cela est apparu pour l'exemple du terrain de football. En dernier lieu, cette limitation pourra résulter de la présence d'autres sites qui entrent en concurrence avec le premier site en s'avérant de meilleurs localisateurs que ce dernier.

Nous avons également pu souligner dans cette section la validité de la notion de cadre. Validité au niveau de la cognition humaine tout d'abord car nous nous transportons sans cesse mentalement d'un endroit à l'autre de l'espace, tout en ne conservant que le minimum d'information nécessaire à la réalisation de notre raisonnement spatial courant. D'autre part, du côté du dialogue homme-machine, et plus spécifiquement du côté du calcul de la référence spatiale, la notion de cadre semble incontournable pour la gestion des ambiguïtés en particulier. Il convient à présent de mieux préciser cette notion.

4.3 Définition et caractérisation de la notion de cadre

4.3.1 Définition générale

Au vu des remarques précédentes, nous proposons la définition suivante de la notion de cadre :

Cadre : entité qui restreint l'univers à une portion d'espace ou à la connexion de portions d'espaces nécessaires et suffisantes, considérée(s) d'un point de vue particulier, autorisant l'achèvement du raisonnement spatial courant.

Remarques :

1. Cette définition de la notion de cadre est en fait une instanciation du concept plus général de contexte que nous défendions au premier chapitre. Pour ainsi dire, le cadre joue le rôle du contexte spatial pertinent.
2. Le fait d'avoir une connexion de portions d'espace est nécessaire pour des exemples du type "*Où est Pierre ?*". En effet, une réponse du type: "*Il n'est pas dans son bureau*" conduira à imaginer l'ensemble des lieux où peut se trouver Pierre, sous la forme d'une réunion de lieux possibles comme "*le couloir*", "*les autres bureaux du bâtiment, etc.*" par opposition à des lieux comme Paris, New-York, etc. Il s'agit en fait de l'ensemble des alternatives potentielles possibles où peut se trouver Pierre. Cette notion d'alternative est très importante comme on le verra par la suite.
3. La notion de point de vue particulier permet de prendre en compte les trois orientations qui peuvent permettre d'orienter un ou plusieurs objets (l'orientation déictique, l'orientation extrinsèque et l'orientation intrinsèque³).

3. Orientation déictique : l'orientation est fondée sur le locuteur.

De manière plus générale, la notion de point de vue particulier est très importante car elle permet de prendre en compte les visions que l'utilisateur a du monde qui l'entoure, en particulier les perceptions qu'il a de ce monde. Jackendoff [1983]⁴ illustre très bien l'importance du locuteur : «*Les choses ne sont rien d'autre que ce que nous en percevons.* », Taylor [1993] insiste également sur ce point : «*L'usage des prépositions spatiales n'est pas uniquement régi par l'objectivisme^(*) mais passe au contraire par la représentation qu'a le locuteur du monde, représentation qui est influencée par nos expériences, nos sensations, les forces agissant sur ces objets* ». La notion de point de vue particulier permet donc également de tenir compte du fait qu'une entité peut tour à tour être considérée comme unidimensionnelle puis bidimensionnelle ou encore tridimensionnelle. Il suffit de se remémorer à ce propos l'exemple d'un étang qui peut être considéré comme bidimensionnel si l'on parle du «*plan d'eau*» ou au contraire comme tridimensionnel si l'on dit : «*l'étang est profond*».

4. Enfin, nous souhaitons insister sur un point qui nous paraît essentiel : la définition de la notion de cadre met en lumière un concept qui nous semble suffisamment général pour pouvoir intervenir au sein de la modélisation de tout type de raisonnement spatial. Toute la difficulté va consister à instancier correctement cette définition et en donner une caractérisation aussi fine que possible. Nous effectuerons cette démarche pour le problème spécifique de la référence spatiale et celui du positionnement dans la suite de ce chapitre.

4.3.2 Les cadres et le problème de la référence spatiale

La notion générale de cadre s'instancie selon la définition suivante pour le problème spécifique de la référence spatiale :

Cadre de référence : entité qui restreint l'univers à une portion d'espace ou à la connexion de portions d'espaces nécessaires et suffisantes considérée(s) sous un point de vue particulier, permettant l'interprétation des expressions référentielles spatiales par différenciation. Cette différenciation opère sur la base d'un choix d'une ou plusieurs entités parmi d'autres entités. Ces autres entités sont dénommées alternatives.

Remarque :

- La notion de différenciation fait écho au modèle que nous avons défini au précédent chapitre qui recherchait la ou les entité(s) sur la base d'une approche contrastive. Nous avons également pu voir dans ce chapitre que le concept sous-jacent aux approches contrastives que nous avons développées n'était autre que la notion d'axiologie. Il semble que cette dernière relève à son tour d'un principe plus général qui est celui d'alternative.

Orientation extrinsèque : l'orientation provient de traits marquants de l'univers comme la gravité.

Orientation intrinsèque : l'orientation découle de l'orientation d'un autre objet.

4. [Jackendoff, 1983] R. Jackendoff : Semantics of spatial expressions, in *Semantics and cognition*, MIT Press, (Cambridge Massachussets), p. 161-187.

4.3.3 Tout est “*alternative*” !

La notion d’alternative constitue l’une des bases de la notion de cadre de référence. Cependant, avant d’illustrer ce point, il nous semble important d’insister sur le côté incontournable de la notion d’alternatives comme support fondamental du fonctionnement de la cognition humaine.

Ainsi, prenons l’exemple de la phrase : “*la mer est belle*”. Il n’y aurait pas grand intérêt à énoncer une telle phrase s’il en était toujours ainsi. Cet énoncé prend en fait toute sa signification au vu des alternatives qu’on peut lui trouver : la mer pouvant être au choix : “*agitée*”, “*pas belle*”, etc. On peut généraliser cette remarque à chaque syntagme de la phrase en remarquant que chacun d’eux met le doigt sur une alternative particulière par opposition à d’autres alternatives. D’autre part, si je dis “*la mer est belle*” ou encore “*la voiture est belle*”, je mentionne des alternatives qui se situent dans l’espace. Je peux également faire état d’alternatives temporelles à l’aide de la phrase “*hier la mer était belle*”. Dans ce cas, cette phrase tire sa signification de l’opposition temporelle entre hier et aujourd’hui, un fort sous-entendu pesant sur le fait que l’état de la mer a changé.

Nous pouvons remarquer que le fait de pouvoir disposer d’alternatives reviendrait en définitive à parler de principe de pertinence et plus particulièrement d’apport minimal d’information à l’interlocuteur. En effet, si une phrase comme “*un être humain ne peut pas traverser les murs*” ne comporte *a priori* pas d’alternatives dans le monde réel, énoncer cette phrase revient à n’apporter aucune information à l’interlocuteur, car elle est toujours vérifiée. Par contre, imaginons que l’on se situe à présent dans une perspective d’apprentissage et que cette phrase soit énoncée à un enfant. On conçoit alors plus aisément qu’on apporte de l’information à ce dernier, l’enfant ayant pu se construire une alternative du type : “*je ne peux pas traverser naturellement les murs mais certaines personnes le peuvent peut-être*”.

Ainsi, plus généralement, il semblerait que quoi que nous énoncions, cela s’entende par rapport à un ensemble d’alternatives. Au moment de produire un énoncé, notre fonctionnement cognitif serait alors le suivant : nous nous plaçons dans un certain ensemble d’alternatives qui est bien souvent implicite, puis nous isolons ou nous nous focalisons sur une ou plusieurs alternative(s) en particulier. De notre point de vue, nous touchons ici à l’un des fondements de la cognition humaine.

Il est à noter que les alternatives bipolaires occupent une place de choix dans la langue en général : “*blanc/noir*”, “*positif/négatif*”, etc. mais aussi plus spécifiquement au niveau spatial : “*droite/gauche*”, “*haut/bas*”, “*près de/loin de*”, etc. Mounin [1972] (déjà cité) va plus loin et conçoit cette structuration binaire comme l’un des caractères fondamentaux du fonctionnement de la pensée humaine.

Pour illustrer ce que pourrait être le fonctionnement d’un système informatique à base d’alternatives, examinons l’exemple de la phrase (123) qui serait effectuée dans le cadre d’un système de multifenêtrage.

Détruis la fenêtre de droite.

(123)

Il serait souhaitable que le système informatique analyse la requête de la manière suivante : on dispose d'objets affichés à l'écran qui sont des fenêtres et leurs alternatives (des icônes par exemple). Le système isole dans un premier temps les fenêtres, dans un second temps, parmi ces dernières, l'une d'elle se situe à droite tandis que les autres (les alternatives) ne se situent pas à droite ou moins à droite que la dite fenêtre. Au terme de son analyse, le système a donc réussi à isoler la bonne fenêtre en éliminant au fur et à mesure les mauvaises alternatives pour se focaliser sur le bon objet recherché.

Comme nous l'avons évoqué dans l'introduction de ce paragraphe, la notion d'alternative semble devoir jouer un rôle essentiel au sein du concept de cadre de référence. Ainsi, imaginons que vous dites à quelqu'un "*Pierre est dans l'église*", en étant vous même et votre interlocuteur dans cette même église. Dans ce cas, la perspective selon laquelle est conçue la scène ne peut pas se réduire au cadre de référence de la figure 4.4.(a). Il faut au contraire, que se trouvent inclus, dans le cadre, l'église et une portion autour de cette dernière afin de supporter l'opposition potentielle : être à l'intérieur ou à l'extérieur de l'église (figure 4.4.(b)), c'est-à-dire l'ensemble des alternatives où peut se trouver Pierre.

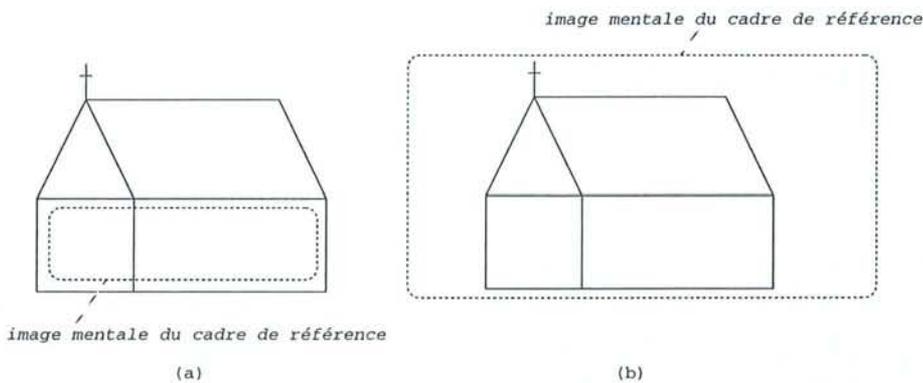


FIG. 4.4 – De la pertinence des cadres de référence.

Nous terminerons ce paragraphe en soulignant que la notion d'alternatives intervient largement dans des disciplines concomitantes à l'informatique. Ainsi, dans le domaine de la créativité, Osborn [1988] insiste sur la nécessité, pour réaliser un bon *brainstorming*⁵, de décomposer tout d'abord le concept ou l'objet que l'on souhaite améliorer en ses composantes élémentaires. On cherchera ensuite pour chacune d'elles le plus d'*alternatives* possibles en différant son jugement concernant la qualité de ces dernières. Enfin, la notion d'alternatives constitue également la pierre angulaire de la notion de *pensée latérale* comme le souligne Edward de Bono [1990] (on trouvera un descriptif détaillé de cette notion à l'annexe B).

5. Littéralement : «tempête des cerveaux». Dans une entreprise, réunion où chacun fournit des idées selon une procédure donnée pour résoudre un problème.

4.3.4 Caractérisation des cadres de référence

Après avoir proposé une définition de la notion de cadre de référence et insisté sur la notion d'alternatives qui lui est sous-jacente, il nous faut à présent caractériser plus finement la notion de cadre de référence. Nous proposons donc les différentes caractéristiques suivantes :

- Le cadre de référence est un concept dynamique, il évolue au fil du discours, par essence même car c'est la portion d'espace minimale (ou réunions de portions d'espace) qui permettra le calcul des référents spatiaux (on verra au paragraphe 4.3.5 l'ensemble de ces évolutions possibles).
- Les frontières du cadre de référence sont relativement floues, même si, bien souvent, elles se précisent au fil du discours. L'étendue du cadre de référence dépend fonctionnellement des objets ou parties d'objets qu'il englobe. Ainsi, par exemple, plus le site est volumineux ou saillant et plus le cadre de référence qui l'entoure est étendu. Par ailleurs, si nous savons que l'étendue du cadre de référence ne doit pas être trop importante, il faut cependant qu'elle soit suffisante, sans quoi, les entités qui s'y trouvent ne prennent pas pleinement leur signification. Par analogie, Langacker [1991] démontre qu'un objet n'a pas de signification, s'il est pris isolément mais qu'il faut nécessairement le placer dans un contexte suffisamment large. Pour illustrer cette idée, il suffit de considérer la figure 4.5 où on ne peut vraiment commencer à parler d'île qu'à partir du "point de vue" 4.5.(d).

Ce point justifie donc la notion de "portions d'espaces nécessaires et suffisantes" que l'on trouve dans la définition des cadres de référence.

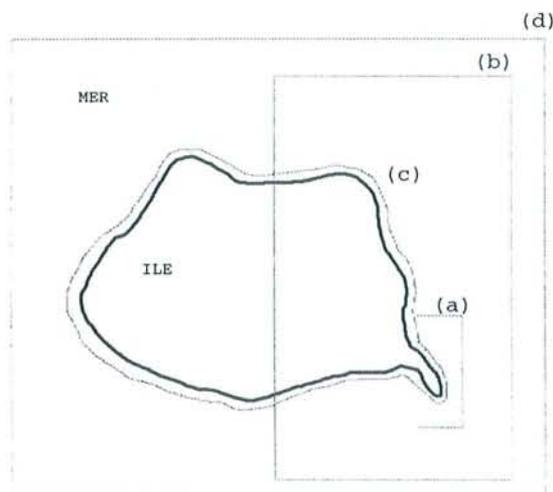


FIG. 4.5 – Différents points de vue sur une île.

- À la notion de cadre de référence, nous adjoignons celle de *focus*. Le *focus* pouvant être défini comme la zone (à l'intérieur du cadre de référence) où se trouve l'entité (respectivement les entités) dont on parle, par opposition aux zones où l'on est sûr qu'elle(s) ne se trouve(nt) pas. Par ailleurs, comme on le verra, lorsque le cadre de

référence reste constant, le focus peut se déplacer à l'intérieur de celui-ci au fil du discours. Le focus peut s'intéresser alors à un élément particulier du cadre, à une zone particulière de ce dernier (le fond, le bord, etc.). Son étendue dépend donc du cadre de référence dans lequel il est inclus mais également de la préposition spatiale mise en jeu. Ainsi, tandis que le cadre de référence regroupe un ensemble d'alternatives, le focus met le doigt sur l'une d'elles en particulier où doit se trouver la ou les cibles recherchée(s). Nous illustrerons cette dernière remarque à l'aide de l'exemple du paragraphe 4.7.

Il convient à présent de situer la notion de focus, que nous emploierons tout au long de ce chapitre, par rapport à la notion de focus employée par Sidner [1986]. Si la notion de focus que nous proposons bénéficie tout comme celle de Sidner du même effet d'attention sur une ou plusieurs entités, en revanche les deux notions sont totalement différentes. Ainsi, tandis que Sidner isole un ou plusieurs objets, de notre côté, nous construisons un cadre de référence puis un focus qui en fait strictement partie. L'(les) entité(s) est(sont) ensuite recherchée(s) à l'intérieur de ce dernier (comme nous le verrons, l'étendue du focus peut éventuellement être agrandie si la(les) cible(s) n'y est(sont) pas trouvée(s)). Il nous semble alors très important que la(les) cible(s) puisse(nt) être "cadrée(s)". En effet, le locuteur ne fait pas que des reprises anaphoriques qui portent sur un ou plusieurs objets, au contraire, il peut très bien faire des reprises de cadre comme cela apparaît dans l'exemple suivant à la figure 4.6 :

<i>Agrandis les carrés de droite.</i>	(124)
<i>Euh non rétrécis le.</i>	(125)
<i>Détruis l'autre.</i>	(126)

Dans ce cas, le suivi discursif du dialogue à l'aide de la notion de cadre de référence (figures 4.6.(b) et 4.6.(c)) permet tout naturellement de prendre en compte la dernière référence "*détruis l'autre*" sur la base du cadre de référence qui englobe les deux carrés (figure 4.6.(d)).

- Dans un cadre de référence donné, on n'a pas nécessairement toutes les orientations (verticalité, latéralité, perspective). En fait, seules les orientations pertinentes sont présentes : dans certains cas une seule direction est pertinente.

Ainsi dans l'exemple des trois gobelets (voir figure 4.7) sous lequel se trouve cachée une pièce dont il faut deviner l'emplacement, les seules directions pertinentes sont la verticalité et la latéralité (la perspective est absente ici).

- Enfin, dans un cadre de référence, toutes les entités qui se trouvent incluses dans son enveloppe spatiale n'en font pas nécessairement partie. Le cadre de référence réalise en fait un filtrage des entités pertinentes. On se rapproche ici de la notion de granularité variable de D. Kayser [1988] et [1992]. À titre d'illustration, imaginons que vous assistiez à un match de football en compagnie d'un ami. Ce dernier énonce alors subitement : "*regarde le numéro 3 ! il vient de frapper le gardien*". Dans ce cas,

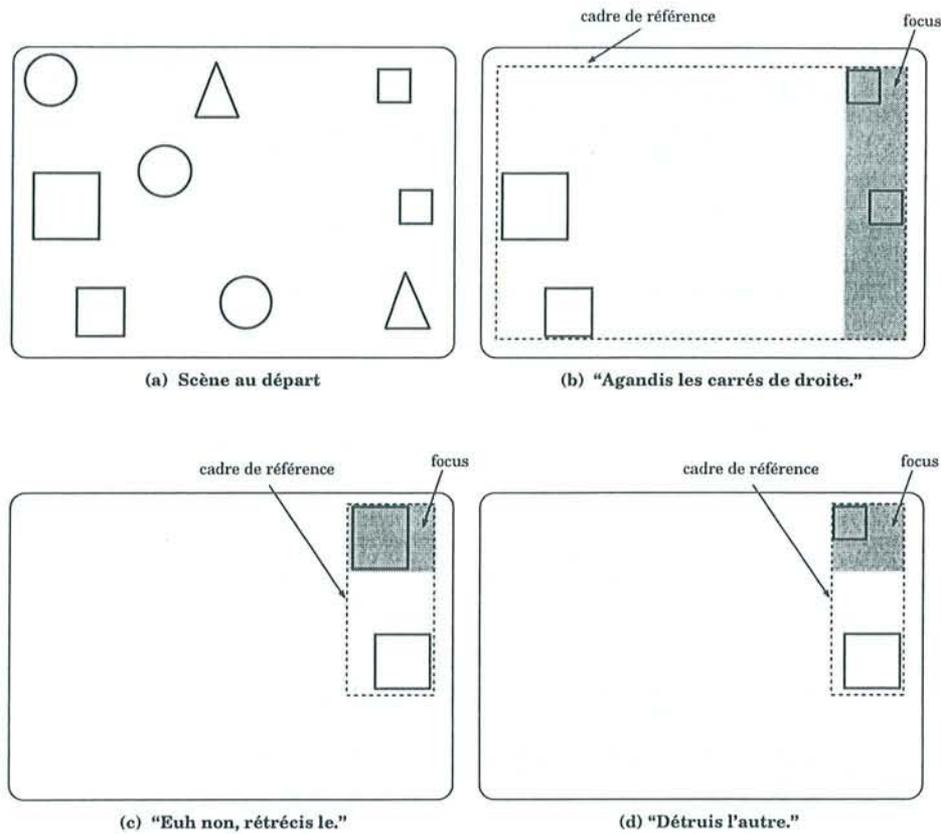


FIG. 4.6 – De l'intérêt des cadres de référence.

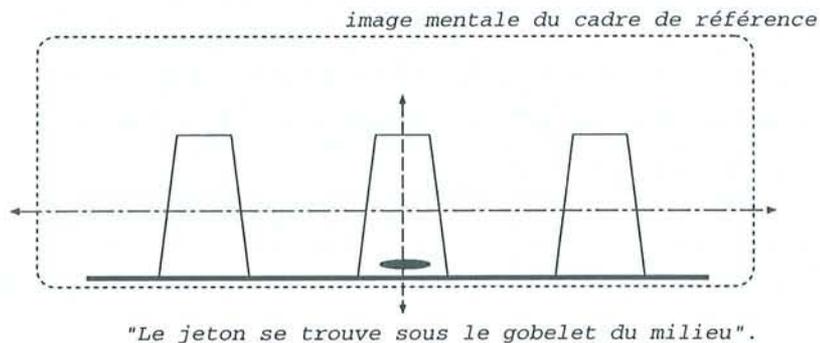


FIG. 4.7 – Deux directions pertinentes: la latéralité et la verticalité.

vous déplacez probablement votre regard et considérez les deux joueurs. Vous vous construisez alors un nouveau cadre de référence qui englobe ces derniers. Il va de soi qu'une partie du gazon est également incluse dans l'enveloppe de ce cadre de référence. Il en va d'ailleurs de même d'une multitude de brins d'herbe. Pour autant, ces derniers ne font pas partie du cadre de référence car ils ne sont pas pertinents.

Par ailleurs, étant donné que le sens d'une préposition n'est pas isolé mais qu'il dépend beaucoup du contexte [Hottenroth, 1993], la notion de cadre de référence est très intéressante car elle permet justement de réduire et de mieux préciser ce

qu'est le contexte.

Nous pouvons résumer les différentes caractéristiques des cadres de référence en remarquant qu'elles se répartissent suivant le triptyque suivant : les spécificités perceptuelles, fonctionnelles et sémantiques. Certaines caractéristiques pourraient faire partie de plusieurs spécificités, nous choisirons cependant de les classer au sein de celles pour lesquelles elles semblent les plus marquantes :

- *Les spécificités perceptuelles :*
 - la notion de point de vue particulier,
 - la granularité de la perception.
- *Les spécificités fonctionnelles :*
 - le choix des orientations mises en jeu (verticalité, latéralité, perspective),
 - l'étendue des cadres de référence qui dépend du site et des entités contiguës.
- *Les spécificités sémantiques :*
 - dynamicité des cadres de référence,
 - étendue suffisante des cadres de référence,
 - focus dépendant de la préposition mise en jeu,
 - la notion d'alternatives.

Voyons à présent quelles sont les possibilités d'évolution des cadres de référence au fil du discours.

4.3.5 Évolution discursive des cadres de référence

Les cadres de référence peuvent évoluer de quatre manières différentes au fil du discours :

- *conservation* : cadre-réf' = cadre-réf⁶. C'est le cas par défaut. Ce cas répond au principe d'économie cognitive, l'utilisateur ne va pas "sauter du coq à l'âne" et va raisonner de manière privilégiée dans le précédent cadre de référence.
- *focalisation* : le nouveau site appartient à l'ancien cadre de référence : cadre-réf' \subset cadre-réf.
- *agrandissement* : cadre-réf' \cap cadre-réf $\neq \emptyset \wedge$ cadre-réf' $\not\subset$ cadre-réf. Ce cas peut se produire lorsqu'une entité du précédent cadre de référence sert à localiser une autre entité (respectivement est localisée par rapport à une autre entité) qui n'appartient pas au précédent cadre de référence.
- *rupture* : cadre-réf' \cap cadre-réf = \emptyset . C'est le cas lorsqu'on localise une entité par rapport à une autre, les deux entités n'appartenant pas au précédent cadre de référence. De notre point de vue, il n'y a jamais de rupture brutale des cadres de référence.

6. cadre-réf désigne l'ancien cadre de référence, cadre-réf' désigne le nouveau cadre de référence qu'il faut calculer.

Il semble qu'une rupture doive plutôt s'assimiler à un agrandissement (figure 4.8, étape 2) puis à une focalisation (figure 4.8, étape 3). Nous illustrerons ce point dans la suite de ce chapitre.

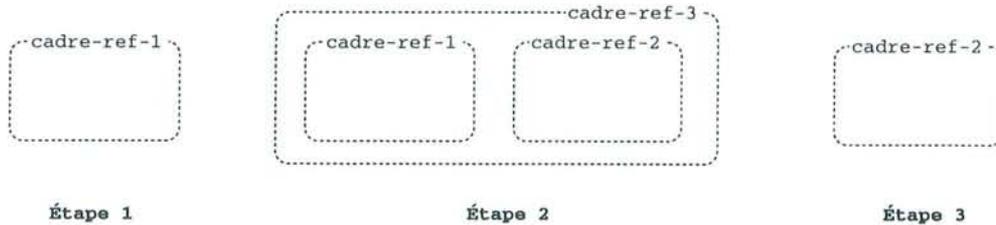


FIG. 4.8 – La rupture au sein des cadres de référence.

Nous restons pour l'instant à un niveau élevé de description des évolutions possibles des cadres de référence. Nous verrons dans la suite de ce chapitre un exemple détaillé, ainsi qu'une formalisation pour le cas spécifique d'un système de multifenêtrage.

Remarques :

1. Le fait que l'être humain raisonne de manière privilégiée dans le précédent cadre de référence peut être validé selon nous par l'expérience de la figure 4.9.(a) tirée de [Rock et Palmer, 1991] où il s'agit de relier les neuf points en traçant quatre segments sans lever le stylo. Les difficultés rencontrées pour effectuer cette tâche (voir la solution à la figure 4.9.(b)) démontrent, selon les auteurs, que les sujets supposent à tort qu'ils ne peuvent pas sortir du carré 3×3 qu'ils dénomment "cadre de référence". Si nous transposons cette expérience pour le problème de l'évolution de nos propres cadres de référence, il semble que cette expérience plaide en faveur d'une certaine inertie de ces derniers au fil du discours ou au fil du dialogue.

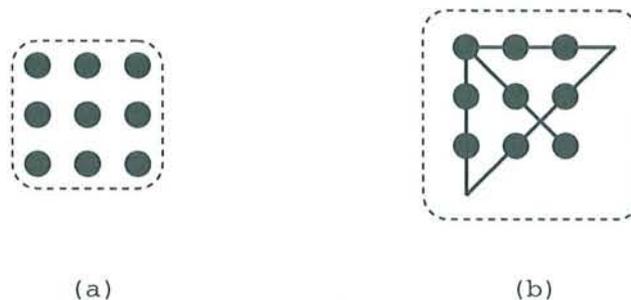


FIG. 4.9 – Raisonnement en priorité dans le précédent cadre de référence.

2. Si l'on présuppose que la notion de cadre de référence reflète certains aspects mis en œuvre par la cognition humaine pour réaliser un raisonnement spatial, il va alors de soi que les quatre changements possibles de cadre (conservation, focalisation, agrandissement et rupture) ont chacun un coût cognitif. Cette assertion nécessiterait bien entendu d'être validée par une expérience psycho-linguistique. Néanmoins, nous

en verrons une illustration pour le cas particulier des énoncés de positionnement qui ont été recueillis au sein du corpus de dialogue homme-machine de Mignot [1995] dans la suite de ce chapitre. En effet, l'opération de rupture semble y apparaître d'ores et déjà, de façon mesurable, comme un temps de latence significatif entre deux phrases qu'énonce le sujet humain.

Nous postulons que le classement des coûts cognitifs pour chacune des évolutions discursives possibles serait le suivant, par ordre de charge cognitive croissante :

- conservation,
- focalisation,
- agrandissement,
- rupture.

La conservation est la moins coûteuse car elle s'inspire du principe d'économie cognitive. La notion de focalisation se fonde en partie sur le précédent cadre de référence en s'intéressant à l'une de ses composantes en particulier, c'est pourquoi elle est plus coûteuse que l'opération de conservation. L'agrandissement oblige le système cognitif à considérer un nouveau cadre de référence partiellement différent du précédent, ce qui rend cette opération sans doute plus coûteuse que celle de focalisation. Enfin, la rupture est la plus coûteuse de ces quatre opérations car elle se fonde sur les opérations d'agrandissement et de focalisation. Nous allons à présent illustrer ces différentes évolutions par un exemple.

Avant de préciser le fonctionnement des cadres de référence pour le cas précis du calcul des référents spatiaux dans un système de dialogue homme-machine, nous allons tenter de valider cette notion en la situant par rapport à certains travaux existants.

4.4 Les cadres de référence et l'existant

4.4.1 Les cadres de référence et l'approche de Vandeloise

Les notions de cible et de site constituent des notions centrales au sein des travaux de Vandeloise. De notre point de vue, pour pouvoir parler de cible et de site, il faut nécessairement se restreindre à un sous-espace, sans quoi nous aurions une multitude de cibles et de sites. Cette restriction apparaît déjà en filigrane dans la notion de "scène objective" du principe de transfert de Vandeloise⁷, mais l'auteur ne développe pas cette notion, or cela nous paraît capital. De même, comment expliquer le fait qu'une cible puisse devenir un site (entre (127) et (128)) et qu'un site puisse devenir une cible (entre (129) et (130))?

Devant la maison_{site} se trouve notre voiture_{cible}. (127)

Il y a des taches_{cible} d'huile sous la voiture_{site}. (128)

7. Revoir en particulier le paragraphe 2.4.2 où nous avons mentionné la notion de *principe de transfert*, que Vandeloise a défini comme suit : le locuteur a la faculté de se déplacer en tout point utile à la perspective selon laquelle il conçoit la *scène objective* qu'il décrit.

Il y a plein de taches_{cible} sur ma voiture_{site}. (129)

C'est parce que tu gares toujours ta voiture_{cible} sous cet arbre_{site}. (130)

Dans notre optique, ces passages *cible-site* et *site-cible* s'expliquent facilement à l'aide des changements de cadre de référence (figure 4.10) : le cadre ayant changé entre 4.10.(a) et 4.10.(b) ainsi qu'entre 4.10.(c) et 4.10.(d), le site (défini comme l'entité la plus saillante dans le cadre de référence) change lui aussi, ainsi que la cible. On pourra remarquer que si l'approche de Vandeloise ne se place pas dans une optique dialogique, elle peut néanmoins prendre en compte les phrases (127), (128), (129) et (130). Ainsi, selon l'auteur, un site présente généralement la caractéristique d'être apparu plus tôt que la cible dans le discours, ce qui permet de prendre en compte les phrases (127) et (128). Pour les phrases (129) et (130) cette règle ne fonctionne bien entendu pas car la voiture, même si elle est apparue avant l'arbre dans le discours, ne joue que le rôle de cible à la phrase (130), l'arbre étant plus massif et facile à distinguer que la voiture.

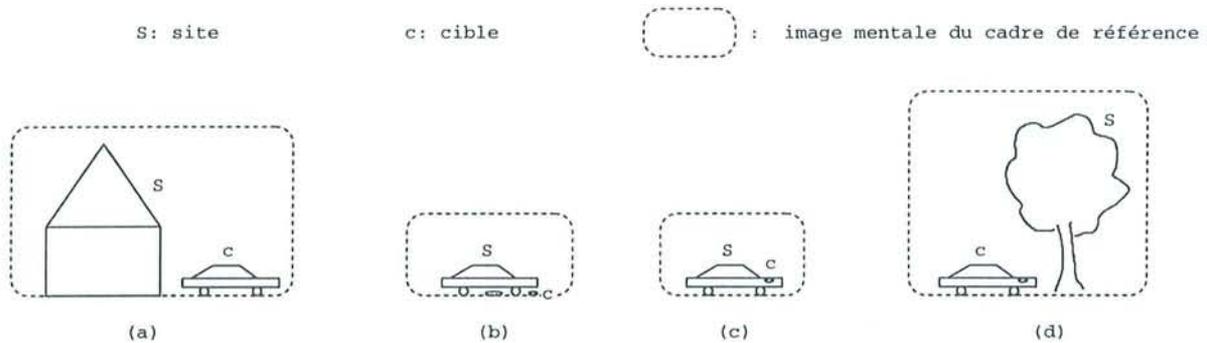


FIG. 4.10 – *Pertinence des cadres de référence.*

Enfin, voici un autre exemple qui semble démontrer que la notion de cadre de référence peut unifier les travaux de Vandeloise. En effet, ce dernier, dans le souci de proposer une modélisation des prépositions “*au-dessus de*” et “*au-dessous de*”, utilise la notion d’axe vertical qu’il caractérise de la façon suivante :

«L’axe vertical est caractérisé par sa direction et son sens. Il est indépendant de la position du locuteur aussi longtemps que le discours n’implique qu’une portion limitée du globe terrestre.» Vandeloise [1986] p. 89-91.

Vandeloise illustre son propos en mentionnant le fait que la verticale française n’est pas la même que la verticale africaine, prenant pour exemple la phrase (131) appliquée à la figure 4.11. De notre côté, nous proposons d’interpréter la notion de verticalité en nous restreignant tour à tour, à telle ou telle portion du globe terrestre : le cadre de référence.

*L’Afrique est en dessous de la France mais le sorcier est
au-dessus de la montagne* (131)

On pourra remarquer que l’exemple (132) ne semble pas tout à fait naturel. Cela provient sans doute du fait que la comparaison s’effectue dans deux cadres de référence

complètement différents, ce qui est beaucoup plus lourd à gérer d'un point de vue cognitif.

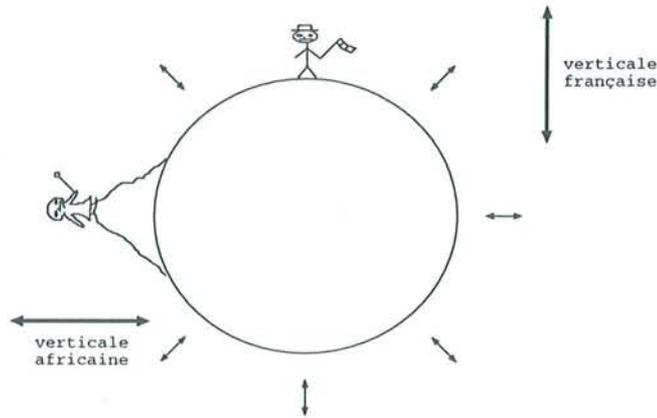


FIG. 4.11 – Deux illustrations de la préposition “au-dessus de”.

4.4.2 Les cadres de référence et l’approche de Asher et Sablayrolles

Pour mémoire, les travaux de Asher et Sablayrolles [1995] réalisent une étude spatio-temporelle des verbes de changement de lieu intransitifs. Si nous occultons les aspects temporels et les aspects liés au prédicat que les cadres de référence ne prennent pour l’instant pas en compte, nous pouvons remarquer que notre notion peut unifier les sept zones de localisation mises en œuvre dans le modèle de Asher et Sablayrolles, car le cadre de référence décrit justement la portion d’espace minimale nécessaire à l’achèvement du raisonnement spatial courant. Par ailleurs, les travaux de Asher et Sablayrolles se situent dans une perspective monologique alors que notre approche s’inscrit dans une perspective dialogique, par la prise en compte du locuteur notamment. Enfin, l’approche de Asher et Sablayrolles décrit des entités en mouvement alors que notre approche opère ses calculs sur des référents dont la position est statique même si leur position peut évoluer au fil du dialogue.

Si le modèle de Asher et Sablayrolles est très abouti, il semble que l’ajout de la notion d’alternatives puisse apporter un plus. En effet, prenons l’exemple de la phrase (132), il semble qu’un fort sous-entendu pèse sur le fait que Jean puisse ne pas courir “*sur place*” mais ailleurs. On assiste à une forte opposition potentielle entre le lieu “*sur place*” et son alternative “*sur place*”. Il semble donc nécessaire d’inclure dans la représentation sémantique de cette phrase une portion d’espace qui entoure le lieu “*sur place*” afin de prendre en compte cette opposition référentielle.

Jean court sur place.

(132)

4.4.3 Les cadres de référence et les aires de recherche de Pribbenow

Nous allons à présent décrire l'approche de Pribbenow [1993] afin de la comparer à la notion de cadre de référence. Pour mémoire, l'approche de Pribbenow réalise l'étude des phrases du type LE REL RO où la LE (*Located Entity*⁸) est localisée par rapport au RO (*Reference Object*⁹) à l'aide d'une préposition REL. Tout comme Bierwish [1983], Pribbenow propose un modèle à deux niveaux :

- 1) un niveau linguistique,
- 2) un niveau conceptuel.

Le premier niveau prend en compte les traits spécifiques du langage tandis que le second se fonde sur un ensemble de concepts (spatiaux) indépendants du langage. Pribbenow se focalise plus spécifiquement sur le second niveau. Chaque expression localisatrice est interprétée en deux étapes :

- La première étape segmente l'espace afin d'obtenir la *primary area*¹⁰ qui contient toutes les localisations possibles de la LE. Voir à ce titre l'exemple de la figure 4.12.(a) pour la phrase (133).

Ma voiture est garée près de la statue. (133)

- Dans un second temps, Pribbenow construit le *search domain*¹¹. Cette dernière aire est obtenue en restreignant la *primary area* aux zones de l'espace où la LE doit se trouver avec la plus grande probabilité. Ces zones de plus grande probabilité ou zones prototypiques s'appellent des *positives priorizations areas*. Leur calcul est réalisé essentiellement à partir de connaissances sur le monde.

Pribbenow restreint ensuite encore davantage les *positives priorizations areas* au moyen des *negatives priorizations areas*. Ces dernières proviennent d'entités entrant en compétition avec le site et restreignent son pouvoir localisateur, voir les figures 4.12.(b) et 4.12.(c).

Au terme de ces deux étapes, Pribbenow obtient le *search domain*. Elle insiste cependant sur le fait qu'au terme de cet algorithme, on n'est plus certain de retrouver la LE. Il se peut, en effet, qu'à un moment donné du calcul, on se soit trop restreint. Pribbenow propose alors de revenir en arrière et d'étendre l'aire de recherche.

8. *Located Entity*: entité localisée. Ce terme n'est pas traduisible par la notion de cible de Vandeloise [1986] car, en parlant d'entités, l'auteur se laisse la possibilité de localiser des situations et des mouvements.

9. *Reference Object*: Vandeloise [1986] parlerait de site.

10. *Primary area*: nous proposons la traduction "aire de recherche au départ".

11. *Search domain*: on pourrait proposer la traduction "aire de recherche affinée" ou "aire secondaire de recherche".

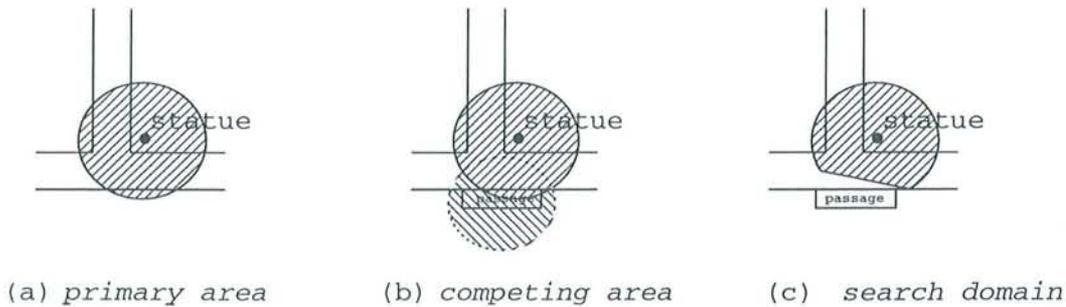


FIG. 4.12 – Construction du “search domain” en considérant les objets environnants.

On relève plusieurs similitudes entre cette approche et la notion de cadre de référence :

- La règle de Pribbenow selon laquelle : «*the bigger the object, the bigger the area*»¹², est prise en compte dans notre approche par le fait que, prototypiquement, le cadre englobant une table aura une extension qui sera plus petite que le cadre englobant une maison, cette extension dépendant du site et de la cible¹³, voir le paragraphe 4.3.4 et [Schang et Romary, 1994a].
- Chez Pribbenow, l’extension de la *primary area* est bien souvent limitée par les objets environnants. De notre côté, le cadre de référence est bien souvent limité par des discontinuités spatio-fonctionnelles, revoir à ce titre le paragraphe 4.3.4.

Parmi, les différences, on notera les deux différences significatives suivantes :

- L’approche de Pribbenow n’est pas discursive.
- Du point précédent, il découle entre autre qu’une séparation nette entre le niveau conceptuel et linguistique telle qu’elle est prônée par Pribbenow n’est pas souhaitable. En effet, cette séparation peut être source de pertes d’indices linguistiques précieux pour prendre en compte le suivi discursif (nous illustrerons cette remarque pour la notion de cadre de référence).

4.4.4 Les cadres de référence et les univers spatiaux de Charolles

Charolles [1996] introduit la notion d’univers spatiaux. Il n’en propose malheureusement pas de définition ni de caractérisation précise. On peut néanmoins définir cette notion par la portion de texte qu’elle indexe. Ainsi, l’expression “*un bourg proche de la frontière brésilienne*” indexe l’ensemble du texte suivant tiré de [Charolles, 1996] :

«*Un bourg proche de la frontière brésilienne. L’annonce de la nationalisation des mines soulève la colère des prospecteurs. Une première manifestation est brutalement réprimée. Un étranger à cheval, Chark, entre en ville et chahute à l’auberge avec une jeune sourde-muette, Maria. Chastin, son père, un prospecteur bigot, la protège tandis que le père Lizzardi, un missionnaire, conseille en vain le calme et la*

12. «*Plus l’objet considéré est grand et plus l’étendue de l’aire de recherche qui l’englobe est grande*».

13. Notre étude porte sur la localisation d’objets et non la localisation de situations ou de mouvements. Remarquons que selon Pribbenow, cela ne porte pas à conséquence car on peut transformer la LE (*located entity*) en un ou plusieurs objets de localisation.

résignation... » (Extrait du résumé du film “La mort en ce jardin” issu d’un magazine de télévision).

Au niveau discursif, un nouvel univers spatial peut soit borner un univers $Us1$ déjà installé (figure 4.13.(a)), soit au contraire l’intégrer (figure 4.13.(b)), les deux univers restant ouverts. Charolles insiste sur le fait que ce sont selon lui les deux seules possibilités que l’on peut rencontrer lorsqu’un univers $Us1$ suit directement un univers $Us2$. Il mentionne, par ailleurs, le fait que le choix entre le bornage et l’intégration se fait sur la base de connaissances extralinguistiques que renferment les introducteurs spatiaux.

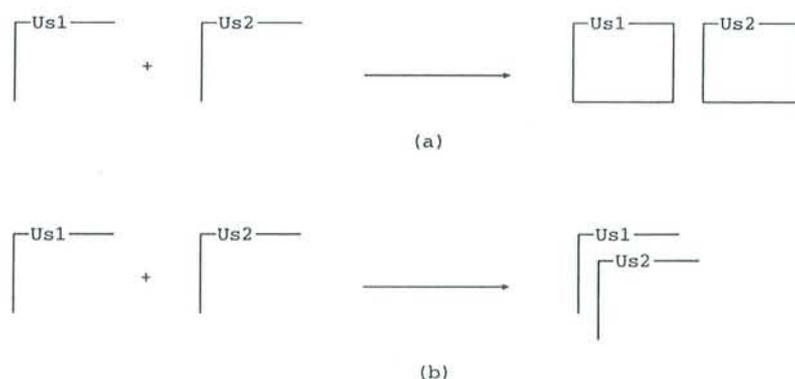


FIG. 4.13 – Les évolutions discursives des univers spatiaux.

On peut faire un parallèle certain entre la notion d’intégration de Charolles et la notion de focalisation que nous proposons au sein des cadres de référence. Il en va de même pour les similitudes entre la notion de bornage et notre notion de rupture. Cependant, comme nous l’avons mentionné dans le paragraphe concernant les évolutions discursives des cadres de référence, une rupture nette d’un cadre de référence nous semble peu naturelle. Ce point n’incommoder pas Charolles qui propose comme exemple de bornage l’exemple suivant :

«*Dans un club de jazz, Greg Wonder, un saxophoniste de renom, dirige un excellent petit orchestre. Sur un terrain d’aviation, Eliot, Vance virevolte aux commandes d’un ULM... »* (Extrait du résumé du film “Attention les dégâts” issu d’un magazine de télévision).

Cet exemple semble effectivement plaider en faveur d’une rupture brutale possible des univers spatiaux comme le montre le passage sans lien apparent de la portion de phrase “*Dans un club de jazz*” à la portion de phrase “*Sur un terrain d’aviation*”. Cependant, en y regardant de plus près, on pourrait sans doute remarquer que ces deux propositions s’inscrivent en fait dans un contexte plus général qui est celui du film. Ce contexte joue alors le rôle de cadre en permettant de donner un sens à ce résumé. Nous plaidons donc en faveur d’une rupture “lente” des univers spatiaux suivant le schéma de la figure 4.8 (p. 139) que nous avons proposé pour les cadres de référence. Dans ce cas, la rupture entre les groupes prépositionnels “*Dans un club de jazz*” et “*Sur un terrain d’aviation*” doivent

plutôt être conçus comme un agrandissement implicite¹⁴ du premier groupe à l'ensemble du film, puis une focalisation de ce dernier au second groupe prépositionnel.

Nous terminerons cette étude en remarquant que la méthodologie de Charolles se complique singulièrement lorsqu'un troisième univers spatial fait son apparition (voir en particulier la figure 4.14).

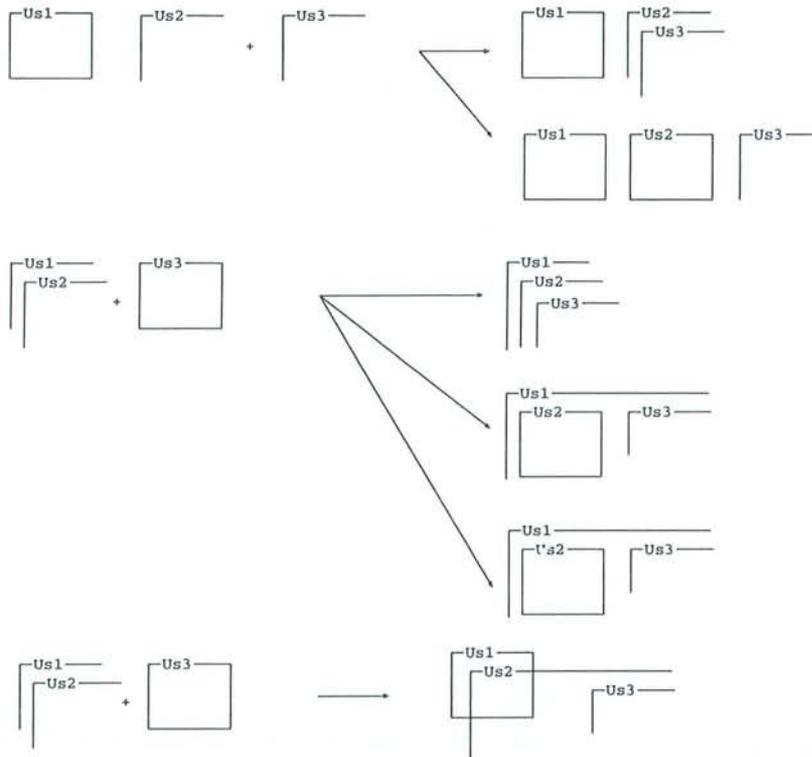


FIG. 4.14 – *Les évolutions discursives des univers spatiaux.*

En conclusion, on peut remarquer que la notion d'univers spatial de Charolles se situe directement dans la lignée de nos préoccupations. Il est regrettable cependant que l'auteur n'en propose pas une caractérisation ou un début de formalisation. Par ailleurs, nous ne partageons pas le point de vue de Charolles selon lequel une rupture brutale des univers spatiaux est possible.

4.4.5 Conclusion

Au terme de cette analyse, il semble que la notion de cadre de référence puisse subsumer certaines composantes des approches que nous venons de mentionner. Le caractère général des cadres de référence nous place ici dans l'opposition entre les modèles généraux et les modèles spécifiques. Tandis que la première classe de modèles manque de précision dans

14. Nous parlons d'implicite ici car le lecteur ne se rend sans doute pas compte de ces opérations mentales.

son comportement, ce qui nécessite une méthodologie d'instanciation, la seconde classe est difficile à étendre.

Voyons à présent, dans une perspective de dialogue homme-machine, quelles sont les composantes à mettre en œuvre afin d'interpréter correctement une requête spatiale de l'utilisateur.

4.5 Connaissances nécessaires à l'interprétation d'une référence spatiale

L'objectif de ce paragraphe est de décrire l'ensemble des connaissances à mettre en œuvre dans un système de dialogue homme-machine en vue de réaliser l'interprétation d'une requête spatiale de l'utilisateur permettant d'aboutir à la désignation d'un ou plusieurs objets dans un système d'affichage graphique. On trouvera à la figure 4.15 un synoptique qui dresse l'ensemble des connaissances sur lesquelles les cadres de référence vont s'appuyer au fil du dialogue.

Le modèle que nous allons présenter ne prend pas en compte les aspects gestuels, quand bien même, il en serait capable moyennant certaines extensions. En effet, comme nous avons pu le souligner dans [Bellalem *et al.*, 1995], un geste de désignation revient, tout comme une référence spatiale purement langagière, à mettre en avant un ou plusieurs objets par opposition à d'autres objets et ce, à l'intérieur d'un cadre de référence donné. L'intégration d'un module de désignation gestuelle fait partie des perspectives de poursuite de ces travaux. Voyons à présent comment construire effectivement le nouveau cadre de référence qui évoluera au fil du dialogue.

4.5.1 Le nouveau cadre de référence

Le modèle des cadres de référence vise à construire pour chaque requête de l'utilisateur le nouveau cadre de référence associé qui permettra de retrouver la ou les cibles recherchée(s) à l'intérieur du focus. Le nouveau cadre de référence est formé d'un 5-uplet (vide au départ) dont il faut instancier chaque item. Les différents items résultent de contraintes provenant de l'analyse syntaxico-sémantique de la requête de l'utilisateur, de l'historique du dialogue, du lexique, etc. (voir la figure 4.15). Le 5-uplet à instancier est le suivant :

1. *prédiction sur les alternatives* possibles et *prédiction d'évolution* du nouveau cadre de référence par rapport à l'ancien cadre de référence,
2. *image mentale* du cadre de référence composée du *type* 2D, $2D\frac{1}{2}$ ou 3D, de son *étendue* et du *focus* qui localise la(les) cible(s),
3. *point de vue* sur le cadre de référence,
4. *contraintes fonctionnelles* entre le(s) site(s) et la(les) cible(s),
5. *provenance de l'image mentale* du nouveau cadre de référence. Ce trait détermine si l'image mentale du nouveau cadre de référence provient d'une conservation, d'une focalisation, d'un agrandissement ou d'une rupture (agrandissement + focalisation) par rapport au précédent cadre de référence.

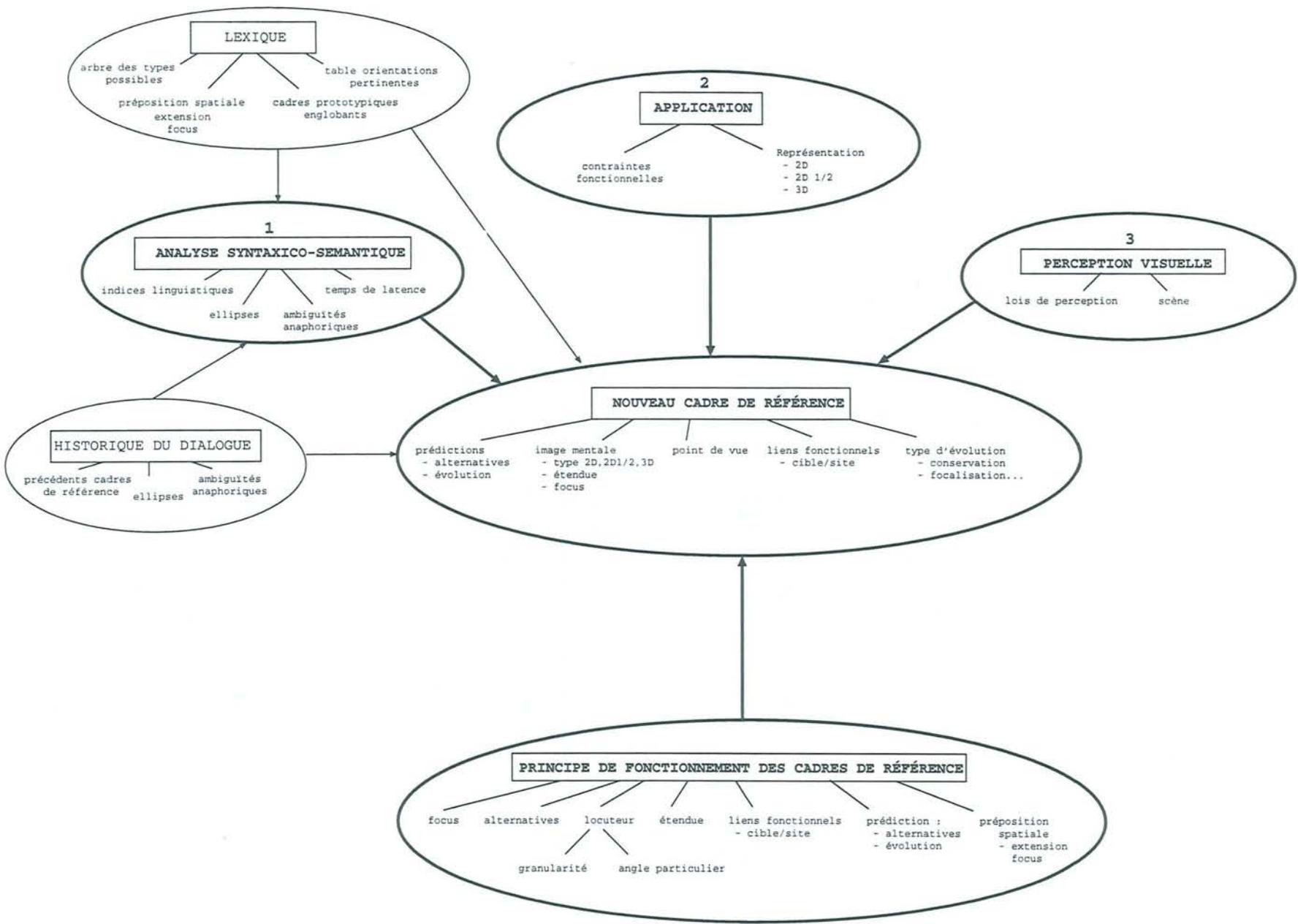


FIG. 4.15 – Connaissances nécessaires à l'interprétation d'une référence spatiale.

Ainsi que nous le verrons dans la suite de ce chapitre, l'élaboration du 5-uplet s'effectue en quatre étapes :

1. l'étape de construction des deux variables *prédiction des alternatives* et *prédiction d'évolution* qui vont nous indiquer respectivement le type des objets contenus dans le nouveau cadre de référence et l'évolution escomptée du nouveau cadre de référence par rapport à l'ancien cadre de référence,
2. l'élaboration des trois champs suivants : image mentale, point de vue et contraintes fonctionnelles,
3. la recherche effective du(des) site(s) et de la(des) cible(s),
4. l'instanciation du champ *provenance de l'image mentale* car on sait à cet instant avec certitude¹⁵ l'opération qui a été effectivement réalisée (conservation, focalisation, agrandissement ou rupture).

Nous verrons au paragraphe 4.6 que la dernière étape de construction du 5-uplet peut s'accompagner de certains affinages du focus. Il suffit en effet de considérer la figure 4.16.(a) pour une requête du type : "rétrécis les rectangles de droite". En s'agrandissant, le focus parvient à isoler directement les rectangles de droite sur la base du groupe homogène perceptif qu'ils forment. Comme cela apparaît à la figure 4.16.(b), le focus ne recouvre qu'une partie des rectangles et n'englobe pas ceux-ci. Cette remarque nécessitera l'accomplissement d'une post-action afin que le nouveau focus englobe toujours les entités référencés une fois qu'il les a localisées. De la même manière, nous insisterons sur le fait que le prédicat peut également modifier l'allure du focus. Il suffit de considérer la requête "agrandis les rectangles de droite" qui nécessitera d'augmenter la taille du nouveau focus une fois l'action du prédicat accomplie (cet agrandissement du focus peut alors avoir également une répercussion sur le cadre de référence, qui englobe nécessairement le focus).

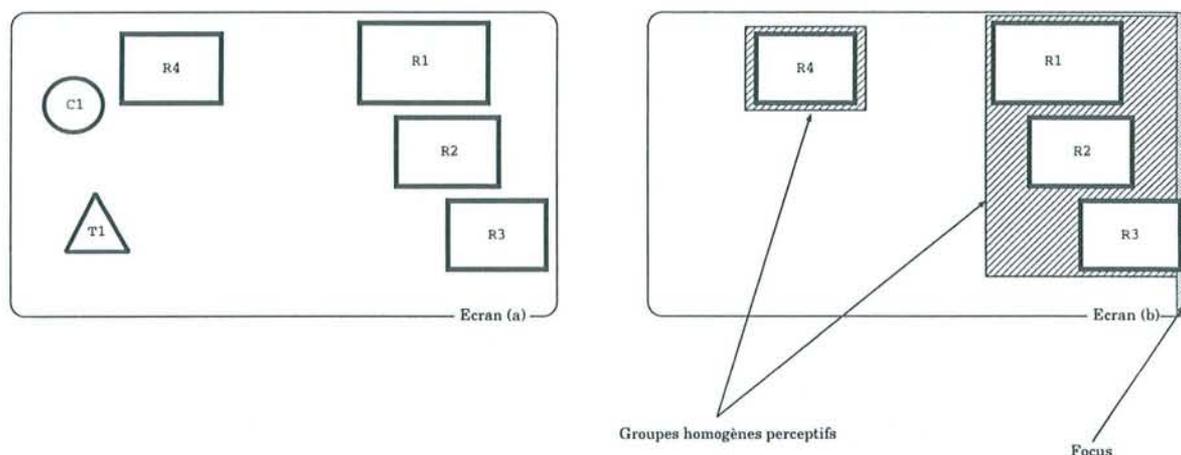


FIG. 4.16 – Inclusion partielle de la cible dans le focus de recherche.

Nous allons décrire à présent en détail la composition de chaque item du 5-uplet.

15. Ce qui n'était pas le cas à la première étape lors du calcul du champ *prédiction d'évolution*.

a) Prédiction des alternatives et prédiction du type d'évolution du cadre de référence

Les deux champs que nous allons présenter à présent relèvent d'hypothèses et ont un rôle important à jouer au moment de l'instanciation du nouveau cadre de référence sur telle ou telle zone de l'écran.

a.1) Prédiction des alternatives

Nous avons pu insister au précédent chapitre sur l'importance du discours. En particulier, certains exemples montraient clairement que l'ensemble où était(étaient) recherché(s) le(s) référent(s) n'était pas indifférent mais qu'il permettait au contraire de nous aiguiller vers la(les) bonne(s) entité(s) recherchée(s). Il suffit de se remémorer l'exemple suivant tiré du paragraphe 3.4.3 :

Déplace les fenêtres vertes. (134)

Mets la fenêtre la plus à gauche en arrière plan. (135)

Nous avons pu conclure alors que la présence du N "*fenêtre*" dans le second énoncé serait inutile s'il s'agissait d'une anaphore : dans l'ensemble des fenêtres vertes, tous les éléments sont des fenêtres ce qui nous ferait plutôt dire "*détruis celle de gauche*" pour le cas d'une anaphore. Au contraire, le bon choix relevait d'une interprétation non anaphorique car l'ensemble dans lequel devrait être recherché le référent était nécessairement constitué de "*fenêtres*" et de "*non fenêtres*" en vertu du schéma de la figure 4.17.(a).

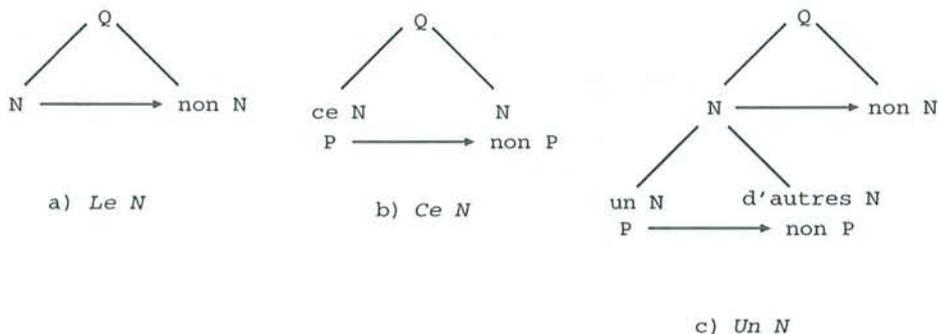


FIG. 4.17 – Les trois schémas d'axiologie selon GaiFFE [92].

Étant donné que ce type d'information est de nature linguistique, c'est au module linguistique que reviendra la charge de préciser quels sont les différentes alternatives possibles à la(les) cible(s) recherchée(s) en termes de type. C'est-à-dire que seront contenues dans le champ "prédiction des alternatives" l'ensemble des types d'objets que l'on peut rencontrer au sein du nouveau cadre de référence lors de son instanciation à l'écran. Ainsi que nous le verrons au paragraphe 4.5.3, le champ "prédiction des alternatives" sera instancié par un module d'analyse syntaxico-sémantique. Ce module reprend et étend l'analyseur syntaxique que nous avons mentionné au précédent chapitre pour y ajouter trois items

(“marqueur”¹⁶, “prédiction des alternatives” et “prédiction d'évolution”):

- (1) *Marqueur*:,
- (2) *Verbe*:,
- (3) *Déterminant de la cible*:,
 - *genre*:,
 - *nombre*:,
 - *type*:,
- (4) *Catégorie de la cible*:,
- (5) *Séparateur*:,
- (6) *Déterminant du site*:,
 - *genre*:,
 - *nombre*:,
 - *type*:,
- (7) *Catégorie du site*:,
- (8) *Préposition spatiale*:,
 - *type de préposition spatiale*:,
- (9) *Type de localisation*:,
- (10) *Prédiction des alternatives*:,
- (11) *Prédiction d'évolution*:

a.2) *Prédiction sur le type d'évolution du cadre de référence*

Nous travaillons actuellement à l'élaboration de *méta-règles* empiriques qui nous renseigneraient sur l'évolution probable du cadre de référence indépendamment de tout autre indice. En effet, on peut prédire empiriquement mais également suivant le principe d'économie cognitive qu'une *rupture* sera rarement suivie d'une autre *rupture*. On peut justifier cette règle par le fait que les actions du locuteur ont forcément une certaine continuité, ce qui implique que l'utilisateur ne passera pas tout son temps “à sauter d'un bout à l'autre de l'écran”. Nous avons déjà pu insister sur le fait qu'une rupture brutale ne nous semblait pas naturelle. Pour autant, il semblerait bien que nous soyons en présence d'un cas de rupture entre la requête (136) et (137) à la figure 4.18:

Détruis le carré qui se trouve à droite du cercle. (136)

Détruis le triangle qui se trouve à gauche du rectangle. (137)

De notre point de vue, il n'y a pas de réelle rupture entre le cadre de référence entourant le cercle et le carré par rapport au cadre de référence entourant le rectangle et le triangle. En effet, si nous nous plaçons dans la peau d'un hypothétique interlocuteur qui aurait à effectuer les requêtes énoncées par le locuteur, il va de soi que la réalisation de la requête (137) passe nécessairement par une reconsidération de la scène affichée sur l'écran dans son ensemble puis d'une focalisation.

L'analyse syntaxico-sémantique de la requête de l'utilisateur peut également fournir des

16. Ce champ sera également détaillé au paragraphe 4.5.3.

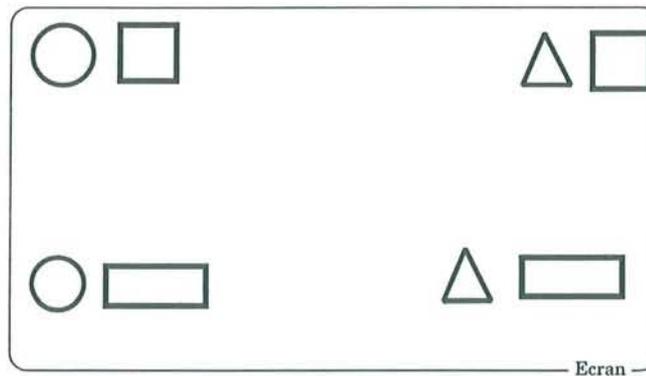


FIG. 4.18 – *Le cas de la rupture au sein des cadres de référence.*

indices linguistiques comme “à présent” ou “maintenant” qui nous conduiront à envisager facilement un *agrandissement* ou une *rupture*. La contribution de ces indices linguistiques sera décrite au paragraphe 4.5.3 concernant l’analyseur syntaxico-sémantique.

b) L’image mentale du cadre de référence

L’image mentale constitue la partie “visuelle” des cadres de référence par opposition aux données propositionnelles incluses également dans le cadre de référence telles que les contraintes fonctionnelles entre le(s) site(s) et la(les) cible(s), la nature de l’évolution réalisée (focalisation, conservation, etc.).

b.1) Type de l’image mentale du cadre de référence

Le type de l’image mentale du cadre de référence (2D, $2D\frac{1}{2}$ ou 3D) est calqué sur le type d’affichage qui est utilisé par l’application.

b.2) L’étendue de l’image mentale du cadre de référence

L’image mentale du cadre de référence matérialise l’étendue effective du cadre de référence dans lequel se trouvera inclus le focus. Excepté le suivi discursif du dialogue qui peut imposer une restriction du cadre de référence à une sous-partie de l’écran, deux concepts inter-dépendants sont de nature à restreindre son étendue afin qu’il n’occupe pas toujours la totalité de l’écran :

- 1) Son étendue dépend du site. En effet, comme nous avons pu le voir au début de ce chapitre, la portée localisatrice du site dépend de sa nature et de sa saillance.
- 2) Les travaux de Pribbenow que nous avons relatés dans ce chapitre démontrent que l’aire de localisation d’un site est limitée par les sites environnants.

1) Pour le premier point qui concerne le lien entre l’étendue du cadre de référence et le site, nous préconisons de stocker ce type d’information dans le lexique qui contiendra pour chaque type d’objet donné, un cadre prototypique englobant, dont la taille est de l’ordre de grandeur de l’objet. Une difficulté subsiste cependant car on ne peut pas toujours

donner *a priori* l'étendue du cadre de référence englobant un objet donné. Pour preuve, il suffit de considérer la figure 4.19.(a) sur laquelle on fait la requête "détruis le cercle qui se trouve au-dessus du triangle". Dans ce cas, l'étendue prototypique du cadre de référence englobant le triangle T4 nous empêcherait d'atteindre le cercle C2. Dans une perspective d'implémentation, la solution que nous retiendrons consistera à agrandir le cadre de référence jusqu'à permettre au focus (qui se trouve strictement inclus dans le cadre de référence) de retrouver effectivement la(les) cible(s) recherchées.

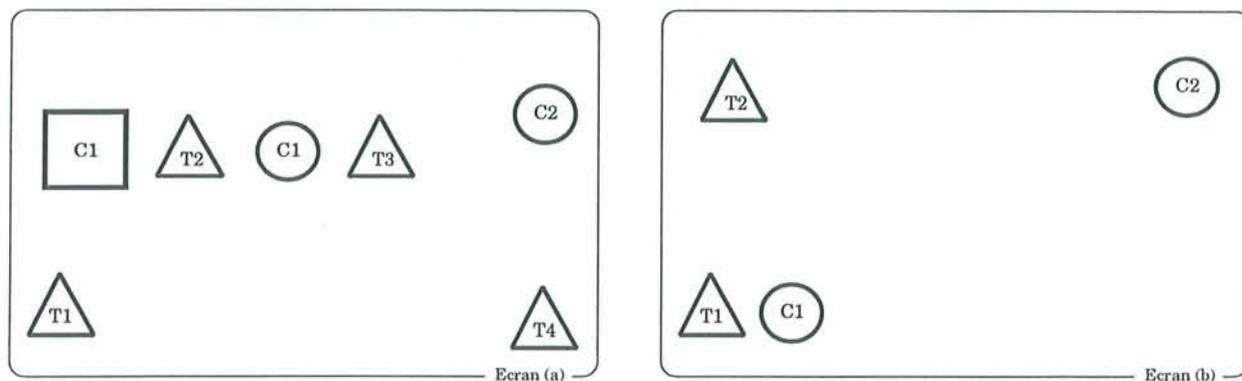


FIG. 4.19 – Le problème de la portée localisatrice limitée du site.

Dès lors, la requête "détruis le cercle qui se trouve au dessus du triangle" sera effectivement comprise par notre approche car l'extension du cadre de référence prototypique associé au triangle T4 englobera le cercle C2 au bout d'un certain temps. Enfin, la requête "détruis le cercle qui se trouve à droite du triangle", appliquée à la figure 4.19.(b), qui présente une ambiguïté sur le site de prime abord, sera là encore résolue par la taille variable du cadre de référence englobant les sites. En effet, le cercle C2 ne fait pas partie du cadre de référence prototypique associé au triangle T2, tandis que le cercle C1, fait partie du cadre de référence prototypique associé au triangle T1 ce qui lève l'ambiguïté.

2) Pour le second problème qui concerne les conflits de localisation entre sites, Pribbenow propose de restreindre l'aire de recherche d'un site donné par les autres sites (revoir la figure 4.12 p. 144). Cependant, cette solution ne semble pas pleinement satisfaisante. En effet, il semble nécessaire dans un premier temps de limiter la portée d'un site donné relativement à d'autres sites uniquement si ces derniers possèdent un degré de saillance semblable ou plus élevé que le site localisateur. À titre d'illustration, il suffit de considérer la figure 4.20 où il paraît difficile de limiter la portée localisatrice de la bouteille par le verre de vin blanc pour une requête du type "donne moi la fourchette qui se trouve à droite de la bouteille", le verre ayant une saillance moins importante que la bouteille. Pour autant, la limitation par des sites de saillance équivalente n'est pas toujours satisfaisante. Il suffit de considérer la figure 4.21 où il est désastreux que le rectangle R2 restreigne la portée localisatrice du triangle T1 n'autorisant plus la requête "détruis le cercle qui se trouve au-dessus du triangle". Il nous paraît donc hasardeux de limiter la portée localisatrice d'un site par un autre.

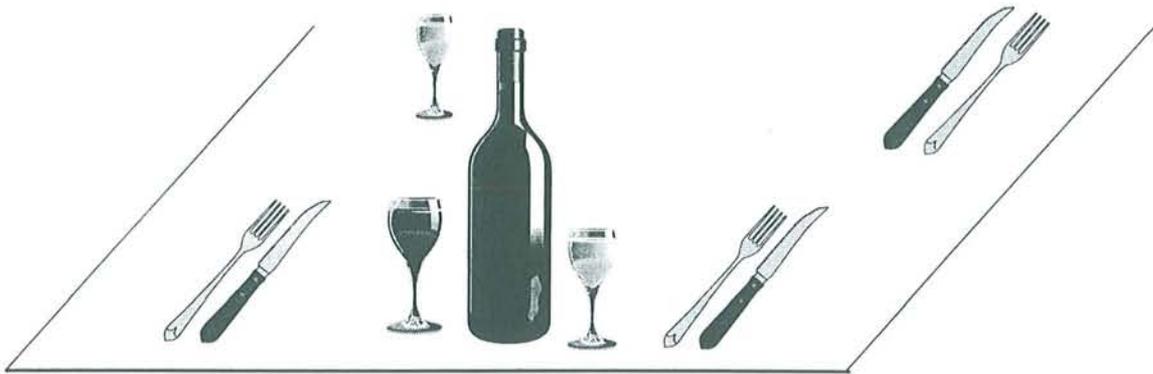


FIG. 4.20 – Exemple de site non obstruant.

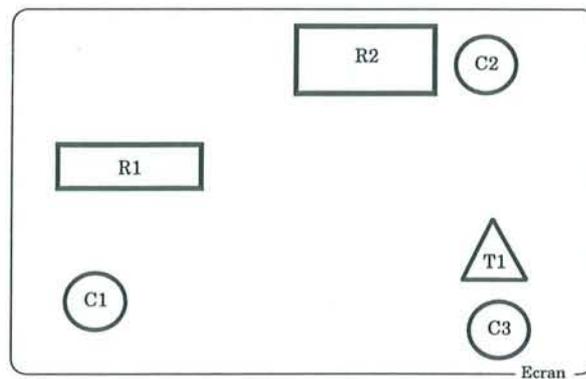


FIG. 4.21 – Limitation de la portée localisatrice d'un site par un autre.

La solution que nous proposons consiste à étendre progressivement le cadre de référence associé à un(ou plusieurs) site(s) donné(s), ce qui permettra de limiter sa(leur) portée localisatrice (le processus d'agrandissement sera détaillé au paragraphe 4.5.2.b). On peut cependant d'ores et déjà remarquer que cette méthode d'agrandissement progressif du cadre de référence permet effectivement de retrouver la(les) cible(s) recherchée(s) lorsque le processus d'extension de recherche du focus n'a pas abouti.

Le rôle de l'image mentale consiste donc à limiter la portée localisatrice du(des) site(s) à l'aide du focus qui se trouve strictement inclus dans l'image mentale. Par ailleurs, l'étendue variable de l'image mentale permet alors de s'adapter aux éventuelles ambiguïtés qui peuvent porter sur la localisation d'une ou plusieurs cibles donnée(s).

b.3) Le focus

Le focus fait partie du cadre de référence et a pour rôle d'isoler la(les) cible(s) référencée(s) par un mécanisme d'extension progressive qui dépend de la préposition spatiale mise en jeu. Alors qu'au précédent chapitre, la recherche de la(les) cible(s) s'effectuait sur la totalité de l'écran, ici, elle ne s'effectue qu'à l'intérieur du cadre de référence dont l'image mentale évolue au fil du discours et n'occupe pas nécessairement la totalité de l'écran.

Ainsi que nous l'avions déjà souligné, l'application a un rôle à jouer dans la construction de l'image mentale du cadre de référence mais aussi dans celle du focus. En effet, si nous nous plaçons en 2D pour la préposition "de droite" par exemple, nous utiliserons l'image mentale et le focus qui apparaissent à la figure 4.22.(a). En revanche, en 3D¹⁷, l'image mentale du focus est en trois dimensions, son agrandissement s'accomplissant à l'aide de plans (figure 4.22.(b)). Bien entendu, nous optimiserons la recherche du carré et de la pyramide pour les requêtes respectives "détruis le carré de droite" et "détruis la pyramide de droite", en restreignant le cadre de référence aux objets qui sont respectivement du type carré et du type pyramide (figure 4.22.(c) et 4.22.(d)). Cette recherche est plus optimale qu'une recherche qui porterait sur la totalité de l'écran et sur la totalité des objets. Les figures 4.22.(a) et 4.22.(b) n'avaient pour unique but que d'illustrer le fonctionnement effectif du focus en 2D et en 3D.

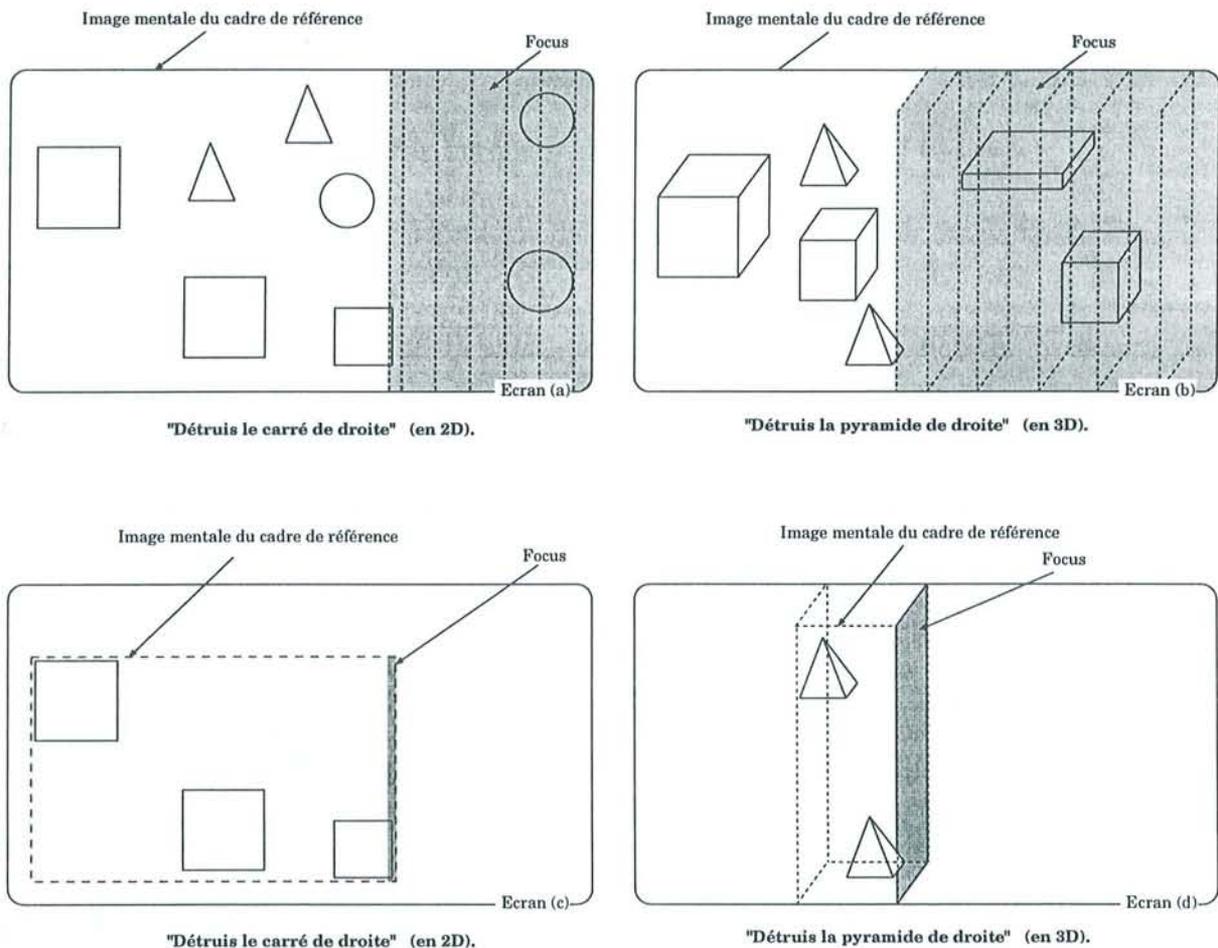


FIG. 4.22 – Focus et extension associés à la préposition "de droite" en 2D et en 3D.

Par ailleurs, il n'y a pas toujours de bijection biunivoque entre un couple (image mentale du focus, extension) et une préposition spatiale. Cela est vrai en particulier pour

¹⁷. Le mode d'affichage en $2D\frac{1}{2}$ étant de type 3D avec peu d'épaisseur, les images mentales élaborées en 3D fonctionnent également en $2D\frac{1}{2}$.

la préposition “sur” dont on dénombre trois usages caractéristiques [Aurnague *et al.*, 1993]. Trois images mentales sont au moins possibles selon les positions relatives de la cible et du site¹⁸ : la cible est plus haute que le site, la cible est au même niveau que le site et enfin, la cible se trouve sous le site (figure 4.23).

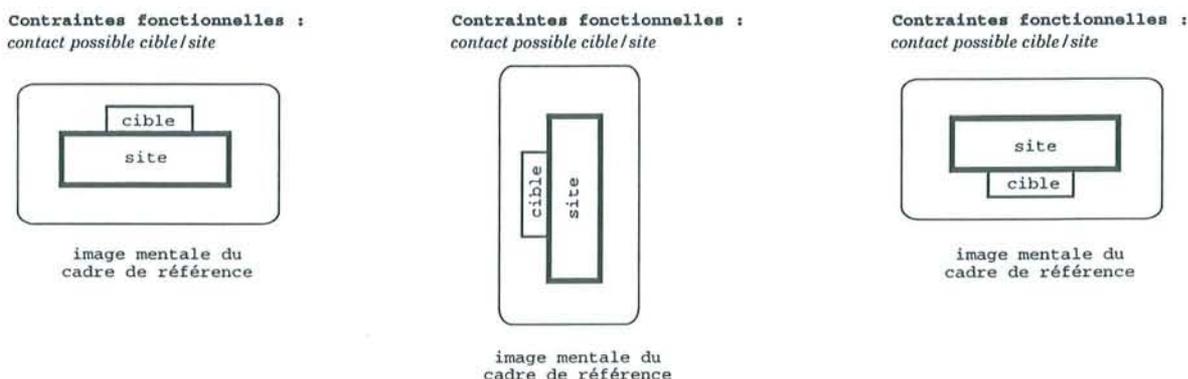


FIG. 4.23 – Différents usages de la préposition “sur”.

Nous détaillerons les contraintes fonctionnelles dans la suite de ce chapitre. Une remarque s’impose cependant à la figure 4.23 concernant ces contraintes. En effet, pour la préposition “sur”, la relation de contact n’est pas toujours vérifiée, il suffit de considérer la figure 4.24 où trois livres sont posés sur une table et un quatrième à terre. On peut alors très bien dire “*donne moi le petit livre qui se trouve sur la table*” pour faire référence au livre B sans qu’il n’y ait pour autant de contact entre ce dernier et la table.

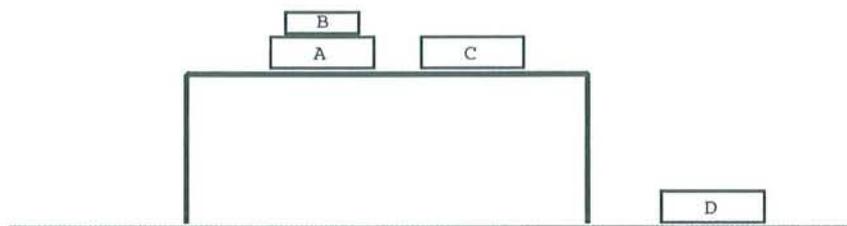


FIG. 4.24 – La préposition “sur” et la relation de contact.

Chaque préposition se voit donc associée une ou plusieurs images mentales du focus et de son extension qui permet(tent) d’isoler une cible ou un groupe homogène perceptif de cibles. Les images mentales des focus et de leurs extensions sont stockées dans le lexique. Nous rappelons à la figure 4.25 l’image mentale du focus et son extension associée correspondant à la préposition “à droite de”.

18. Il conviendrait bien entendu de fournir les images mentales pour le cas où plusieurs cibles seraient localisées relativement à un site (138) et le cas où plusieurs cibles seraient localisées relativement à plusieurs sites (139) :

Les stylos se trouvent sur le bureau. (138)

Il n’y a que des verres et des fourchettes sur les tables! (139)

(la dernière phrase étant énoncée dans un restaurant par exemple).

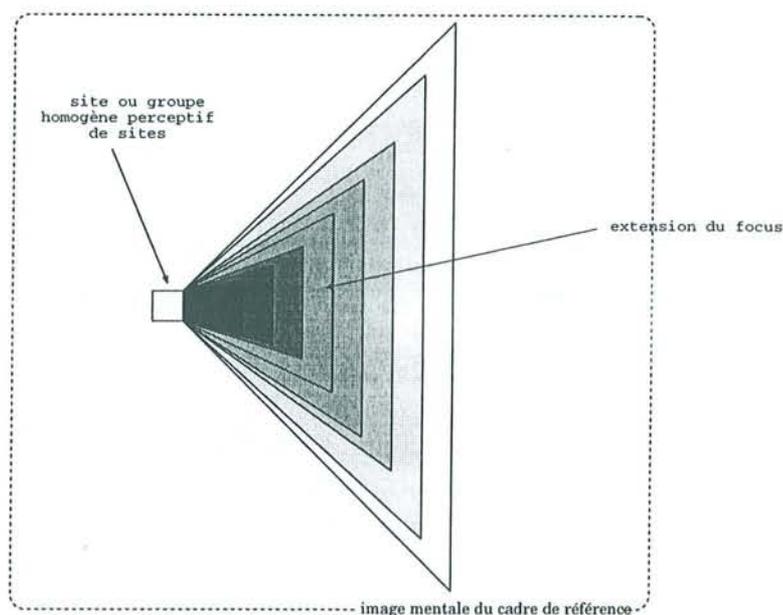


FIG. 4.25 – Image mentale du focus et extension associés de la préposition “à droite de”.

Jusqu'à présent, nous avons décrit les images mentales des focus de recherche associés aux prépositions “à droite de”, “à droite”, “de droite”, “à gauche de”, “à gauche”, “de gauche”, “en haut”, “d'en haut”, “au-dessus de”, “en bas”, “d'en bas”, “en dessous de”, “en haut à droite”, “en haut à droite de” “en bas à gauche”, “en bas à gauche de”. Nous travaillons à l'élaboration des images mentales associées aux prépositions “devant”, “de devant”, “derrière”, “de derrière”, “dans” et “près de”. Pour les prépositions “devant”/“derrière”, la notion de masquage aura sans doute un rôle très important à jouer au sein des images mentales élaborées. Pour la préposition “près de”, l'enveloppe du(des) site(s) sera déterminante pour la localisation de la(les) cibles, ce qui imposera au focus de coller finement au(x) site(s). Enfin, pour la préposition “dans”, tout comme la préposition “sur”, plusieurs images mentales seront associées, les aspects fonctionnels ayant un rôle décisif à jouer (on pourra se reporter à [Vieu, 1991] pour un descriptif détaillé des aspects fonctionnels concernant cette préposition).

Nous rappelons le principe de fonctionnement de la recherche de la(des) cibles : une fois le cadre de référence construit, le focus va s'étendre progressivement jusqu'à localiser une cible (pour le cas de référence au singulier) ou un groupe homogène perceptif (pour le cas de la référence plurielle). Le processus complet de recherche avec la gestion des éventuelles ambiguïtés portant sur le(s) site(s) sera décrit en détail au paragraphe 4.6.

c) Le point de vue

Le point de vue gère trois aspects. Le premier procède des orientations qui sont présentes au sein du cadre de référence. Ces dernières découlent directement de la préposition spatiale intervenant dans la requête de l'utilisateur. Nous proposons le tableau suivant d'association entre chaque préposition spatiale et les orientations pertinentes mises en jeu :

<i>Préposition spatiale</i>	<i>Orientations pertinentes</i>
“à droite de”, “à droite”, “de droite” “à gauche de”, “à gauche”, “de gauche”	latéralité latéralité
“en haut à droite”, “en haut à droite de”.....“en bas à gauche”, “en bas à droite de”	latéralité et verticalité latéralité et verticalité

Il convient alors d’instancier l’(les) orientation(s) pertinente(s) sur le(s) site(s) en gé- rant en particulier les conflits entre les orientations déictique, extrinsèque et intrinsèque. L’emplacement du focus dépend alors crucialement de l’orientation choisie (voir en parti- culier la figure 4.26.(a) où la partie en pointillés désigne le devant intrinsèque (s’il existe) d’un objet par opposition à la figure 4.26.(b) qui relève exclusivement de l’orientation déictique. Nous ne poursuivrons pas plus en avant cette voie dans le cadre de cette thèse et reportons le lecteur aux travaux précédemment cités concernant les conflits possibles entre ces différentes orientations.

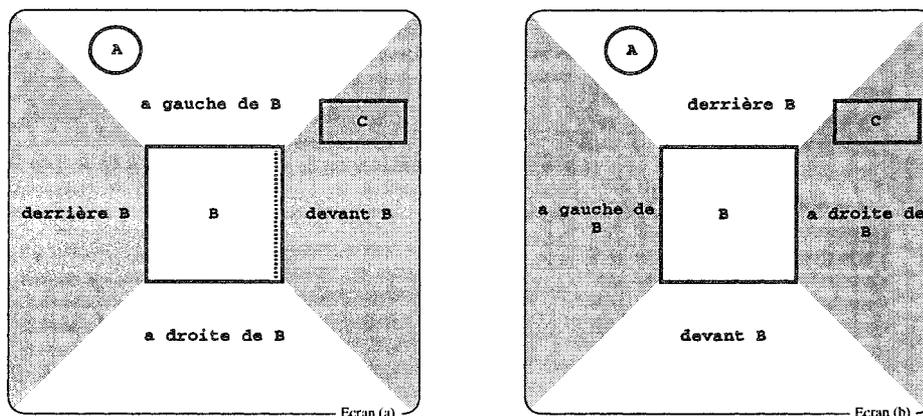


FIG. 4.26 – *Différentes orientations intrinsèques.*

Le second aspect concerne la conception que l'utilisateur a de ses objets. Deux points sont ici déterminants. Tout d'abord, l'utilisateur peut concevoir les entités du monde tantôt de manière linéaire ou planaire ou au contraire de manière tri-dimensionnelle (il suffit de reprendre l'exemple de l'étang du paragraphe 4.3.1 qui pouvait être considéré suivant plusieurs dimensions). Il en découle que l'utilisateur peut donner différentes dénominations aux entités auxquelles il souhaite faire référence. La difficulté concernant les différentes dénominations qui peuvent être utilisées par le locuteur sera résolue en ayant recours à un arbre des types possibles qui fait partie du lexique et que nous détaillerons au paragraphe 4.5.3.

Le troisième aspect concerne les capacités de l'utilisateur à structurer perceptivement la scène perçue. On dénombre trois facultés essentielles sur lesquelles nous reviendrons plus longuement au paragraphe 4.5.4 qui sont respectivement la capacité à percevoir des groupes d'objets, la capacité à percevoir différents niveaux de détail et enfin la capacité à produire un filtre visuel.

d) Le type d'évolution

Ce trait du cadre de référence indique simplement la provenance du nouveau cadre de référence par rapport à l'ancien : s'agit-il d'une conservation, d'une focalisation, d'un agrandissement ou d'une rupture ? Ce choix sera effectué sur la base de caractérisations fournies au paragraphe 4.3.5 (p. 138). On choisira alors entre les quatre types d'évolutions possibles selon les inclusions éventuelles du nouveau cadre de référence par rapport à l'ancien cadre de référence.

e) Contraintes fonctionnelles cible/site

Nous nous limiterons à la considération des contraintes fonctionnelles de contact et d'inclusion¹⁹. La relation de contact permet de guider la recherche de la cible en indiquant que le site touche la cible. Il incombera ensuite au focus de recherche de localiser précisément l'objet suivant l'image mentale mise en jeu.

Pour les relations d'inclusion, qui interviendront essentiellement au sein des requêtes faisant intervenir la préposition “dans”, nous distinguerons l'inclusion partielle (exemple à la phrase (140)) de l'inclusion totale. En ce qui concerne l'inclusion totale, nous distinguerons l'inclusion totale classique *inclusion* (x,y) (exemple à la phrase (141)) de l'inclusion du type partie-tout *composant-assemblage* (x,y) (exemple à la phrase (142)). Tandis que la première indique simplement que la cible x se trouve à l'intérieur du site y , la seconde indique que x fait partie de y et implique que x ne peut pas être placé ailleurs qu'à l'intérieur de y . Dans ce cas, la partie a une fonctionnalité par rapport au tout et/ou elle a des frontières bien déterminées, enfin le tout n'est pas considéré comme une collection dont la partie serait un élément. On pourra là encore se reporter aux travaux de Vieu [1991] pour trouver une taxinomie des fonctionnalités associées aux différentes inclusions possibles.

La rose est dans le vase. (140)

La voiture est dans le garage. (141)

La pédale fait partie du vélo. (142)

L'ensemble de ces contraintes sont stockées dans l'application et nous verrons au paragraphe 4.5.4 qu'elles peuvent permettre de structurer perceptivement la scène affichée.

Il semble à présent nécessaire de détailler les différentes contributions qui vont permettre de construire effectivement ce 5-uplet.

4.5.2 L'historique du dialogue et le lexique

a) L'historique du dialogue

Outre son rôle dans la résolution des éventuelles ellipses et ambiguïtés référentielles telles qu'elles seront détaillées au paragraphe 4.5.3, ce module a la charge de stocker

19. Même si nous n'avons pas encore proposé d'images mentales pour les prépositions “sur” et “dans”.

les précédents cadres de référence. Néanmoins, il faut tenir compte du caractère plus ou moins dynamique de la tâche visualisée, notamment, lorsque celle-ci peut évoluer indépendamment des cadres du locuteur (si on imagine par exemple le cas d'un système d'affichage d'une console de trafic aérien). Dans ce cas, seuls les tous derniers cadres sont susceptibles d'être pertinents et il sera nécessaire d'introduire un mécanisme d'oubli. À l'inverse, pour une tâche qui n'évolue que par le fait des ordres de l'utilisateur, des références portant sur des cadres de référence beaucoup plus anciens seront possibles.

b) Le lexique

Notre propos n'est pas ici de décrire l'ensemble des informations que doit renfermer un lexique (on trouvera un bon descriptif des problèmes attachés à l'élaboration d'un lexique dans [Bouyer, 1997]). Au contraire, nous nous focaliserons uniquement sur les connaissances qu'il faut inclure dans le lexique afin de calculer correctement les références spatiales.

b.1) Focus associé(s) à une préposition spatiale

Ainsi que nous avons pu l'indiquer ci-dessus concernant l'image mentale du cadre de référence, chaque préposition spatiale se voit assignée un focus d'extension²⁰. Ce focus doit permettre d'isoler une ou plusieurs cibles sur la base de propriétés spatiales que possèdent les objets recherchés et que ne possèdent pas les autres objets. Chaque couple (préposition spatiale, extension du focus) est donc stocké dans le lexique.

b.2) L'arbre des types possibles

Comme nous l'avons esquissé au précédent chapitre, si l'application donne *a priori* un type aux objets, l'utilisateur peut malgré tout utiliser un nom plus générique que celui associé au type de l'objet. Par exemple, si on étiquette un objet de l'application comme un "triangle", l'utilisateur peut quand même l'appeler une "figure". Autrement dit, il faut pouvoir détecter qu'un "triangle" est une "figure".

Pour gérer cette situation, il est donc nécessaire de répertorier au niveau du lexique l'ensemble des termes pouvant se rapporter à chaque type des objets de l'application. Le nom du type de l'objet, qui constitue le terme privilégié de l'application pour dénommer les objets, peut alors être considéré comme un prototype à partir duquel sont dérivés tous les autres termes susceptibles d'être employés par l'utilisateur.

b.3) Table des orientations pertinentes

Le lexique contient également la table des orientations pertinentes qui indique l'ensemble des orientations associées à chaque préposition spatiale donnée (revoir en

20. Pour les prépositions "dans" et "sur", plusieurs focus seront possibles puisqu'il n'y a pas de bijection biunivoque entre ces prépositions et leur image mentale associée.

particulier le tableau de la page 158).

b.4) Cadres prototypiques englobants

Deux cas peuvent se produire dans une tâche donnée : soit les objets manipulés sont de nature relativement similaire (du fait en particulier de leurs fonctionnalités associées), soit au contraire, leur nature est hétérogène et leur saillance n'est pas équivalente. Lorsque les objets ne possèdent pas une saillance équivalente, il convient de ranger pour chacun d'eux dans le lexique, un cadre de référence prototypique associé.

Dans le cas contraire, on peut calculer pour chaque objet son cadre prototypique englobant suivant un algorithme identique. Nous proposons de construire le cadre de référence prototypique sur la base du rectangle englobant l'objet graphique en agrandissant ce dernier d'une distance d verticalement et horizontalement (figure 4.27). La distance d est définie comme la longueur du plus petit côté du rectangle englobant l'objet considéré. Nous avons choisi cette valeur afin d'éviter que le cadre de référence prototypique n'ait une étendue trop importante (voir le triangle de la figure 4.27). Enfin, dans le cas où le site est en fait constitué de plusieurs sites, on réalise la même opération mais à partir du rectangle englobant les sites concernés (voir les cercles à la figure 4.27).

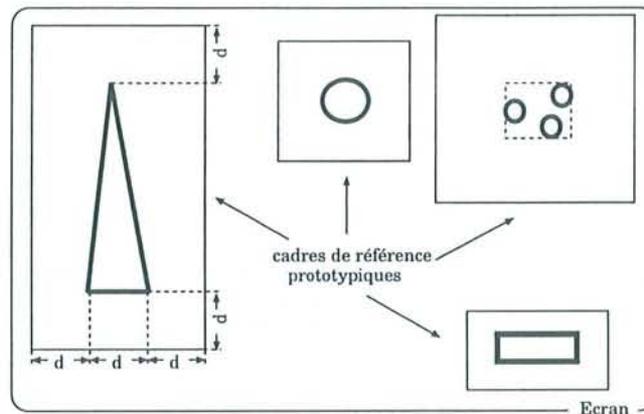


FIG. 4.27 – Construction du cadre prototypique associé à un ou plusieurs sites.

On pourrait sans doute discuter le choix portant sur l'allure de l'enveloppe du cadre de référence prototypique englobant le(s) site(s). En effet, pourquoi choisir plutôt une enveloppe de type rectangulaire plutôt que circulaire par exemple ? Par ailleurs, on peut très bien imaginer des scènes où se dégagent naturellement des enveloppes englobantes non rectangulaires. Ce fait apparaît à la figure 4.28.(a) où on a bien envie de regrouper perceptivement les triangles de gauche sous la forme d'un grand triangle (figure 4.28.(b)).

Notre choix a porté sur une enveloppe englobante de type rectangulaire car un fort caractère anthropomorphique pèse sur une telle enveloppe²¹. En effet, à la différence d'un cercle, un rectangle possède naturellement les orientations de verticalité et de latéralité. Enfin, pour les contextes aussi marqués que celui de la figure 4.28, qui nécessiteraient le

21. Nous avons pu insister au second chapitre sur l'importance déterminante des caractères anthropomorphiques dans l'appréhension de l'espace qui est faite par l'homme.

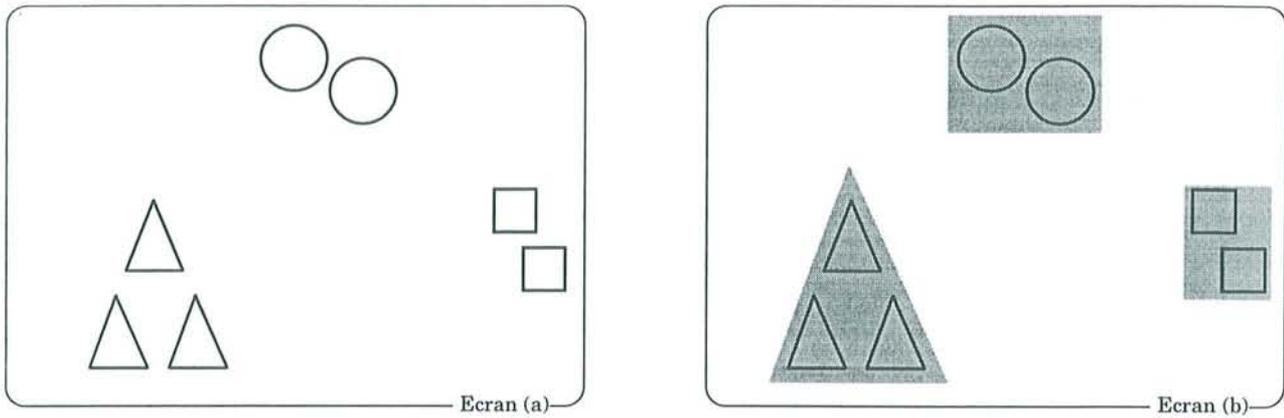


FIG. 4.28 – Le problème des enveloppes englobantes.

recours à des enveloppes triangulaires, notre choix s’est porté sur les enveloppes rectangulaires car la détection d’un triangle englobant à la figure 4.28 nécessiterait d’appréhender finement la scène affichée.

4.5.3 Analyse syntaxico-sémantique

a) Introduction

Cette composante est chargée de gérer les requêtes de l’utilisateur en langage naturel. Elle doit permettre d’analyser la phrase de ce dernier en produisant en sortie, lorsqu’ils existent, le prédicat, la catégorie de la cible, la catégorie du site, leurs déterminants associés, la ou les préposition(s) spatiale(s), et un éventuel marqueur linguistique (il peut s’agir par exemple de “à présent”, “maintenant”, etc. qui peut nous renseigner sur l’évolution probable du cadre de référence). Pour analyser les requêtes de l’utilisateur, nous supposons que nous disposons d’un analyseur syntaxico-sémantique qui nous renvoie une représentation de la phrase traitée. Dès lors, si nous nous préoccupons d’analyser les phrases du type <marqueur>* <verbe> <déterminant cible>* <cible> <séparateur>* <préposition spatiale> <déterminant site>* <site>*²², qui recouvrent bon nombre des expressions référentielles purement spatiales, l’analyseur syntaxico-sémantique nous renverra alors le 11-uplet associé. Par exemple, pour la phrase “À présent, détruis le carré qui se trouve à droite du triangle”, l’analyseur syntaxico-sémantique nous renverra le 11-uplet suivant :

- (1) Marqueur : “À présent”,
- (2) Verbe : “détruire”,
- (3) Déterminant de la cible : “le”,
 - genre : masculin,
 - nombre : singulier,
 - type : défini,
- (4) Catégorie de la cible : “carré”²³,

22. L’étoile : “*” indique que l’élément associé peut être présent ou non.

23. Pour les catégories de la cible et du site, nous nous limiterons aux catégories cercle, carré, triangle, rectangle, cube, sphère, disque, pyramide et parallélépipède.

- (5) Séparateur : “*qui se trouve*”,
- (6) Déterminant du site : “*le*”,
 - genre : masculin,
 - nombre : singulier,
 - type : défini,
- (7) Catégorie du site : “*triangle*”,
- (8) Préposition spatiale : “*à droite de*”,
 - type de préposition spatiale : simple,
- (9) Type de localisation : relative.
- (10) Prédiction des alternatives : $\neq\{C\}$ et $\neq\{S\}$,
- (11) Prédiction d'évolution : conservation.

Voici donc quelques exemples de phrases que l'analyseur syntaxique peut analyser :

- Agrandis les cubes de droite.* (143)
- Détruis celui de gauche.* (144)
- Agrandis les disques d'en bas.* (145)
- Efface ceux de droite.* (146)

- À présent, efface la sphère qui se trouve au-dessus de la pyramide.* (147)

Il nous faut à présent détailler la construction des champs “catégorie de la cible”, “catégorie du site”, “prédiction des alternatives” ainsi que les éventuelles ellipses et ambiguïtés anaphoriques²⁴.

b) La catégorie de la cible et du site

Nous avons pu insister sur le fait qu'étant donné que l'utilisateur, pour faire référence à un objet, n'utilise pas nécessairement le type qui lui est associé dans l'application, il est nécessaire d'avoir recours à un arbre des types possibles pour renvoyer à l'analyseur syntaxico-sémantique la bonne catégorie associée au site et à la cible.

c) Le champ “prédiction des alternatives”

Cet item fournit une prédiction sur l'ensemble des types d'objets que l'on devrait rencontrer au sein du nouveau cadre de référence. Cette prédiction est conditionnelle car il peut advenir quelle ne soit pas vérifiée. À titre d'illustration, il suffit de reconsidérer l'exemple de la requête “*détruis le carré de droite*” qui serait formulée sur l'écran de la figure 4.29.(a) qui ne contient que des carrés.

Dans ce cas, le champ “prédiction des alternatives” (ainsi que nous allons le décrire dans la suite) sera instancié par la valeur “*carré*” \cup “*carré*”. Bien entendu, aucun “*¬carré*” ne se trouve affiché à l'écran ce qui rend impossible l'instanciation du nouveau cadre de référence

24. La phrase (149) comporte une ambiguïté anaphorique sur le type de l'objet à détruire :

- Agrandis les carrés de droite.* (148)
- Détruis celui de gauche.* (149)

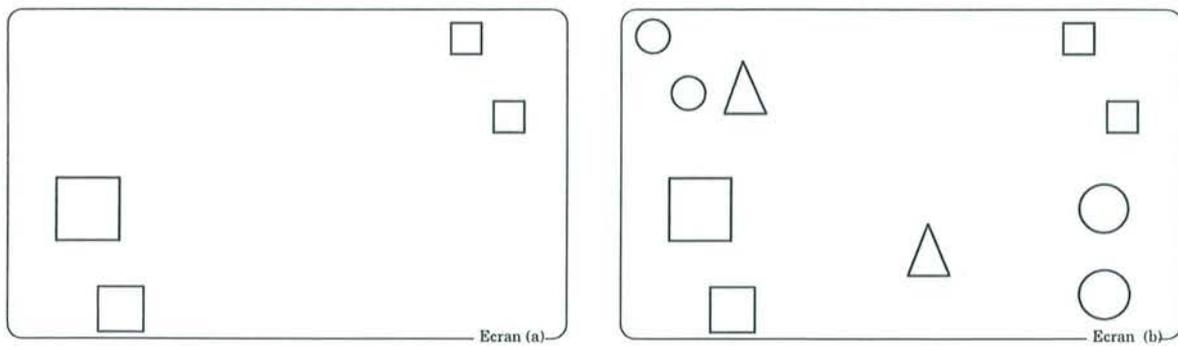


FIG. 4.29 – Une application manipulant des figures géométriques.

sur une zone de l'écran et même sur la totalité de ce dernier. Cependant, comme nous l'avons fait remarquer au paragraphe 3.4.3, le fait de parler de \neg “carré” est légitime car on pourrait aussi bien mentionner l'écran, le fond de l'écran, etc. Il ressort de cet exemple que les cas où aucune zone de l'écran ne correspond au champ “prédiction d'évolution” imposeront au nouveau cadre de référence d'être instancié sur la totalité de l'écran.

D'autre part, il est important de souligner que cette prédiction s'applique au moment de l'instanciation du nouveau cadre de référence sur l'écran car ce dernier subit ensuite différentes étapes d'affinage. À titre d'illustration, il suffit de considérer la requête “*détruis le carré de droite*” qui serait formulée cette fois sur l'écran de la figure 4.29.(b) qui contient des objets du type \neg “carré”, à savoir, des “cercles” et des “triangles”. Dans un premier temps, le cadre de référence est instancié sur une zone de l'écran qui contient des “carrés” et des \neg “carrés”. Dans un second temps (ainsi que cela apparaîtra au paragraphe 4.6), le nouveau cadre de référence sera filtré pour ne contenir que des carrés ce qui permettra dans un second temps d'isoler le carré recherché.

Dans la suite, nous considérerons que $\{C\}$ désigne le type de la(des) cible(s) et que $\{S\}$ désigne le type du(des) site(s). Ainsi que nous allons le décrire à présent, le champ “prédiction des alternatives” peut prendre trois valeurs différentes: $\{C\}$, $\{C\} \cup \neg\{C\}$ et enfin, une valeur dont nous savons juste qu'elle est différente de $\{C\}$ et différente de $\{S\}$. Ainsi que cela apparaîtra dans la suite, le choix entre ces différentes valeurs s'effectuera sur la base de critères syntaxiques. Voici un exemple pour chacune de ces possibilités (étant donné que les requêtes en “*celui de*”, “*celle de*” et “*ceux de*” reposent nécessairement sur une phrase précédente. Dès que ce sera nécessaire, nous donnerons à chaque fois cette phrase sachant que le champ “prédiction d'évolution” que nous indiquons concerne la seconde requête):

Requête	Champ "prédiction des alternatives"
"Agrandis les carrés de droite."	$\{C\} \cup \neg\{C\}$
"Agrandis les carrés de droite." "Détruis celui de gauche."	$\{C\}$
"Agrandis les cercles de gauche." "Détruis celui qui se trouve à gauche du triangle."	$\neq\{C\}$ et $\neq\{S\}$
"Détruis le cercle qui se trouve à droite du carré."	$\neq\{C\}$ et $\neq\{S\}$

Une première remarque s'impose sur ce tableau concernant la valeur $\{C\} \cup \neg\{C\}$ (qui est instanciée par la valeur "carré" \cup "carré" pour la phrase "Agrandis les carrés de droite."). Mathématiquement, l'ensemble $\{C\} \cup \neg\{C\}$ devrait désigner l'ensemble de tous les types possibles car tout objet est soit du même type que la cible, soit il ne l'est pas, et dans ce cas, il fait partie des $\neg\{C\}$. De notre point de vue, l'ensemble $\neg\{C\}$ ne désigne en aucun cas l'ensemble de tous les objets restants. Pour s'en convaincre, imaginons que vous vous trouviez dans une salle à manger lors d'un repas. Si la maîtresse de maison vous dit "passe moi le sel", il va de soi que les objets qui s'opposent dans ce cas au sel, c'est-à-dire les "sel" ne sont en aucun cas, votre chaise ou la nappe de la table par exemple. On peut imaginer au contraire que les "sel" sont par exemple constitués du poivre, de la moutarde, etc.

La seconde remarque découle de la précédente: comment calculer effectivement les $\neg\{C\}$? (le calcul que nous allons décrire pour les $\neg\{C\}$ s'applique également à celui des $\neg\{S\}$). Pour calculer les objets du type $\neg\{C\}$ qui s'opposent aux objets du type $\{C\}$, nous proposons de recourir à l'arbre des types possibles (qui se trouve stocké dans le lexique) et de remonter d'un cran dans cet arbre pour accéder à l'hyperonyme^(*) d'un type $\{C\}$ donné. Dans ce cas, les $\neg\{C\}$ seront déterminés par l'ensemble des objets qui ne sont pas du type $\{C\}$ tout en faisant partie de l'ensemble hyperonyme du type $\{C\}$. À titre d'illustration, prenons le cas d'une tâche de manipulation de figures géométriques. Dans ce cas, le type "triangle" a pour hyperonyme l'ensemble des figures géométriques et l'ensemble des $\neg\{C\}$ est formé par l'ensemble des objets du type "cercle", "rectangle", etc.

Voyons à présent sur quels critères syntaxiques il convient de faire le choix entre les valeurs $\{C\}$, $\{C\} \cup \neg\{C\}$ et la valeur: $\neq\{C\}$ et $\neq\{S\}$ ²⁵. Nous rappelons que la structure syntaxique des phrases que nous nous proposons d'étudier est de la forme: <marqueur>* <verbe> <déterminant cible>* <cible> <séparateur>* <préposition spatiale> <déterminant site>* <site>*. Voici un tableau qui dresse l'ensemble des valeurs du champ "prédiction des alternatives" en fonction de la syntaxe de la requête de l'utilisateur ("celui de" \vee "celle de" \vee "ceux de" indique un choix parmi ces trois démonstratifs):

25. Ainsi que cela sera explicité dans la suite, cette valeur indique que le nouveau cadre de référence ne pourra pas être uniquement constitué d'objets du même type que la cible, pas plus qu'il ne pourra être uniquement constitué d'objets du type du site au moment de son instanciation sur l'écran.

<i>Structure syntaxique de la requête</i>	<i>Champ “prédiction des alternatives”</i>
<marqueur>* <verbe> <déterminant -démonstratif> <séparateur>* <cible> <préposition spatiale>	$\{C\} \cup \neg\{C\}$
<marqueur>* <verbe> “celui de” ∨ “celle de” ∨ “ceux de” <séparateur>* <prép. spatiale>	$\{C\}$
<marqueur>* <verbe> <déterminant cible> <cible> <séparateur> <préposition spatiale> <déterminant site> <site>	$\neq\{C\}$ et $\neq\{S\}$

Nous allons à présent justifier la valeur de chacun de ces champs.

- La structure syntaxique de la phrase est du type <marqueur>* <verbe> <déterminant non démonstratif> <séparateur>* <cible> <préposition spatiale> et la valeur du champ “prédiction des alternatives” est $\{C\} \cup \neg\{C\}$. Les déterminants démonstratifs ont été écartés afin d’éviter les requêtes du type : “*détruis celle de gauche*” qui seront traitées au point suivant. Les requêtes que nous nous proposons d’étudier recherchent nécessairement leur(s) référent(s) dans un ensemble formé d’objets du type $\{C\} \cup \neg\{C\}$ en vertu des schémas 4.17.(a) et 4.17.(c) 4.17 (page 150).

Examinons à présent en détail un exemple pour illustrer ce fait : considérons la requête (150) qui serait formulée à la figure 4.30.(a).

Agrandis les carrés de droite. (150)

La requête (150) fonctionne en deux étapes : on isole tout d’abord les objets qui sont du type “carré” par opposition à ceux qui ne sont pas de ce type. Dans un second temps, parmi les objets du type “carré”, on isole ceux qui se trouvent à gauche. Ces deux étapes sont facilement représentables par le formalisme de Gaiffe selon la figure 4.31. Dans ce cas, l’ensemble initial Q dans lequel sont recherchés les référents est constitué de “carrés” et de \neg “carrés”.

Par ailleurs, on peut représenter ces différentes étapes de recherche à l’aide de la notion de cadre de référence (figure 4.30). Dans un premier temps, le nouveau cadre de référence s’instancie sur une zone de l’écran qui contient des “carrés” et des \neg “carrés” (figure 4.30.(b)). Dans un second temps, un filtrage s’effectue pour ne retenir que les objets du type “carré” (figure 4.30.(c)). Enfin, dans un dernier temps, une structuration perceptive de la scène s’opère afin d’isoler les carrés qui se trouvent effectivement à droite (figure 4.30.(d)).

Il est important de noter que le fait de prédire quels sont les types qui font partie du nouveau cadre de référence permet de comprendre certaines phrases qui seraient ambiguës dans l’absolu. Il suffit de considérer la suite de phrases (151) et (152) qui seraient formulées à la figure 4.32.

Agrandis les cercles de droite. (151)

Détruis le cercle de gauche. (152)

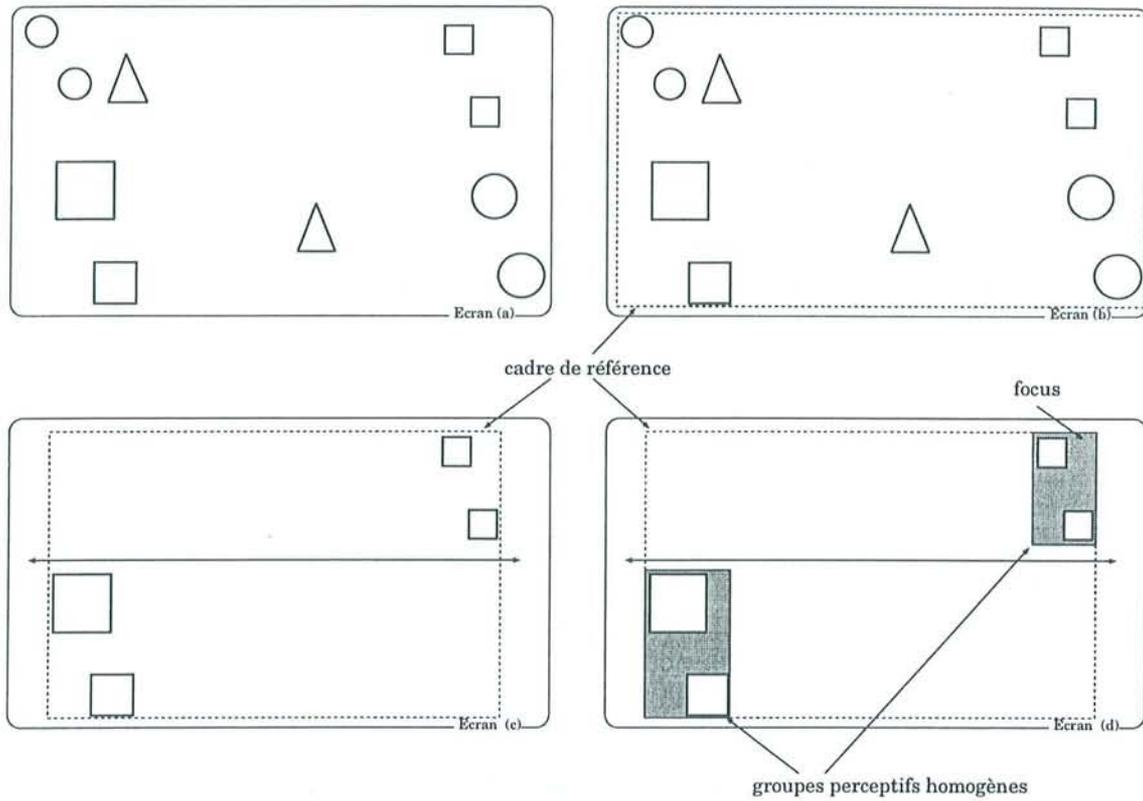


FIG. 4.30 – Instanciation du nouveau cadre de référence et affinage.

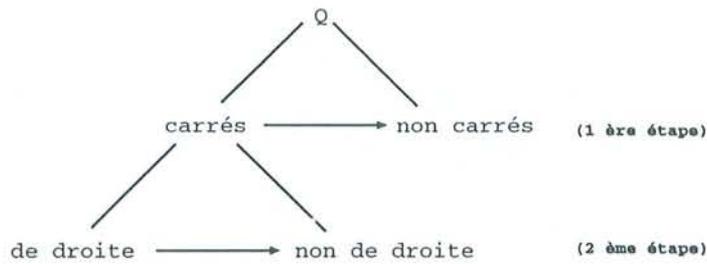


FIG. 4.31 – Représentation du champ “prédiction des alternatives” défini par $\{C\} \cup \neg\{C\}$ à l'aide de la notion d'axiologie.

Notre modèle prédirait une instanciation du nouveau cadre de référence dans un ensemble qui contiendrait des objets du type “cercle” et \neg “cercle” ce qui permet en particulier de rejeter le cercle numéro 2. Ce rejet paraît naturel car on peut espérer que si l'utilisateur avait souhaité détruire le cercle numéro 2, il aurait plutôt utilisé la requête “détruis celui de gauche” qui évite de répéter le mot “cercle”, ce qui place cette phrase plus en continuité de la requête (151).

- La structure syntaxique de la phrase est du type <marqueur>* <verbe> “celui de” <séparateur>* <prép. spatiale> et la valeur du champ “prédiction des alternatives”

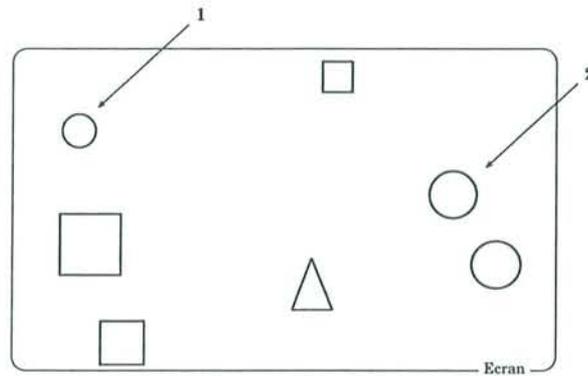


FIG. 4.32 – Une levée d’ambiguïté sur la base du champ “prédiction des alternatives”.

est {C}. Ces requêtes recherchent nécessairement leur(s) référent(s) dans un ensemble qui ne contient que des objets du même type {C} que la(les) cible(s) en vertu du schéma 4.17.(b) que l’on peut étendre aux démonstratifs en “celui de”, “celle de” et “ceux de”.

Examinons à présent en détail, pour la figure 4.33.(a), la requête (153) qui fait suite à la phrase “Agrandis les carrés de droite.”.

Détruis celui de gauche.

(153)

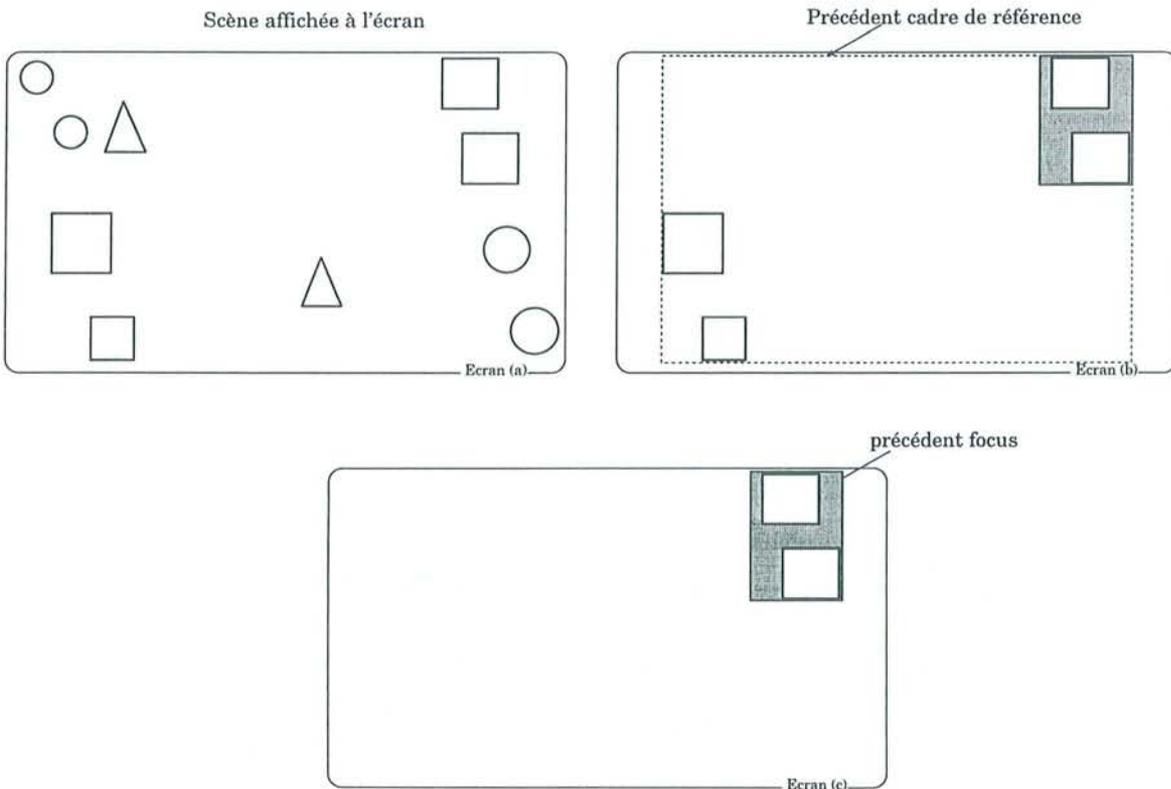


FIG. 4.33 – Une scène représentant des figures géométriques.

Deux choix sont alors possibles pour l'instanciation du nouveau cadre de référence associé à la phrase (153) : dans l'ancien cadre de référence (figure 4.33.(b)) ou dans le précédent focus (figure 4.33.(c)). On assiste ici à un réel cas d'ambiguïté. Une brève réflexion pourrait cependant nous conduire à retenir plutôt le carré qui vient d'être agrandi car il vient précisément de subir une modification ce qui le rend plus saillant. Au contraire, on pourrait préférer les carrés isolés de gauche car ils préservent le point de vue projeté sur la scène par la phrase "*agrandis les carrés de droite*" qui oppose les carrés de gauche à ceux de droite. De notre point de vue, cet exemple relève d'une réelle ambiguïté que parvient à détecter le modèle proposé (il convient dans ce cas d'interroger l'utilisateur afin qu'il se prononce plutôt en faveur de tel ou tel carré).

Il est à noter que dans certains cas, la structuration perceptive de la scène peut permettre d'effectuer le choix entre une référence portant dans le précédent cadre de référence ou au sein du précédent focus. Il suffit de considérer la requête (154) qui porte également sur la figure 4.33.(a).

Détruis ceux de gauche.

(154)

Sur cet exemple, il paraît indéniable que cette requête fait référence aux carrés qui se situent complètement à gauche car aucun sous-groupe perceptif homogène ne se dégage des carrés qui se situent à droite ce qui interdit l'instanciation du nouveau cadre de référence au sein du précédent focus.

- Les requêtes restantes en $\langle \text{marqueur} \rangle^* \langle \text{verbe} \rangle \langle \text{déterminant cible} \rangle \langle \text{cible} \rangle \langle \text{séparateur} \rangle \langle \text{préposition spatiale} \rangle \langle \text{déterminant site} \rangle \langle \text{site} \rangle$ ne peuvent pas s'instancier dans un ensemble qui serait uniquement constitué d'objets du type $\{C\}$ ni dans un ensemble qui serait uniquement constitué d'objets du type $\{S\}$. Cela découle de la nature des phrases étudiées, qui font intervenir un ou plusieurs site(s) en plus d'une ou plusieurs cible(s). Il en ressort qu'au moins deux types d'objets seront présents au sein du cadre de référence²⁶.

Il convient à présent d'expliquer pourquoi on ne peut pas faire de prédiction plus précise sur la nature du type des objets qui seront présents au sein du nouveau cadre de référence lors de son instanciation. Considérons la phrase : "*détruis le carré qui se trouve à droite du cercle*", qui fait intervenir des syntagmes en "*le N*"²⁷. Dans ce cas, selon le schéma de Gaiffe [1992], on pourrait penser que la nature des objets contenus dans le nouveau cadre de référence seraient du type $\{C\} \cup \neg\{C\} \cup \{S\} \cup \neg\{S\}$. Cela est vérifié à la figure 4.34.(a) ainsi qu'à la figure 4.34.(b) car un \neg "carré" peut être un "cercle" et réciproquement.

En revanche, on peut très bien trouver des exemples qui montrent que la prédiction $\{C\} \cup \neg\{C\} \cup \{S\} \cup \neg\{S\}$ n'est pas toujours exacte. Ainsi, considérons un logiciel qui pour-

26. Cette remarque n'est bien sûr pas vérifiée pour certaines phrases dont la structure syntaxique est différente de $\langle \text{marqueur} \rangle^* \langle \text{verbe} \rangle \langle \text{déterminant cible} \rangle^* \langle \text{cible} \rangle \langle \text{séparateur} \rangle^* \langle \text{préposition spatiale} \rangle \langle \text{déterminant site} \rangle^* \langle \text{site} \rangle^*$. Il suffit en effet de considérer la requête "*détruis le petit cercle qui se trouve à droite du grand*" qui pourrait très bien s'instancier dans un ensemble qui ne contient que des cercles.

27. La préposition "*du*" étant la contraction de la préposition "*de*" et de l'article "*le*".

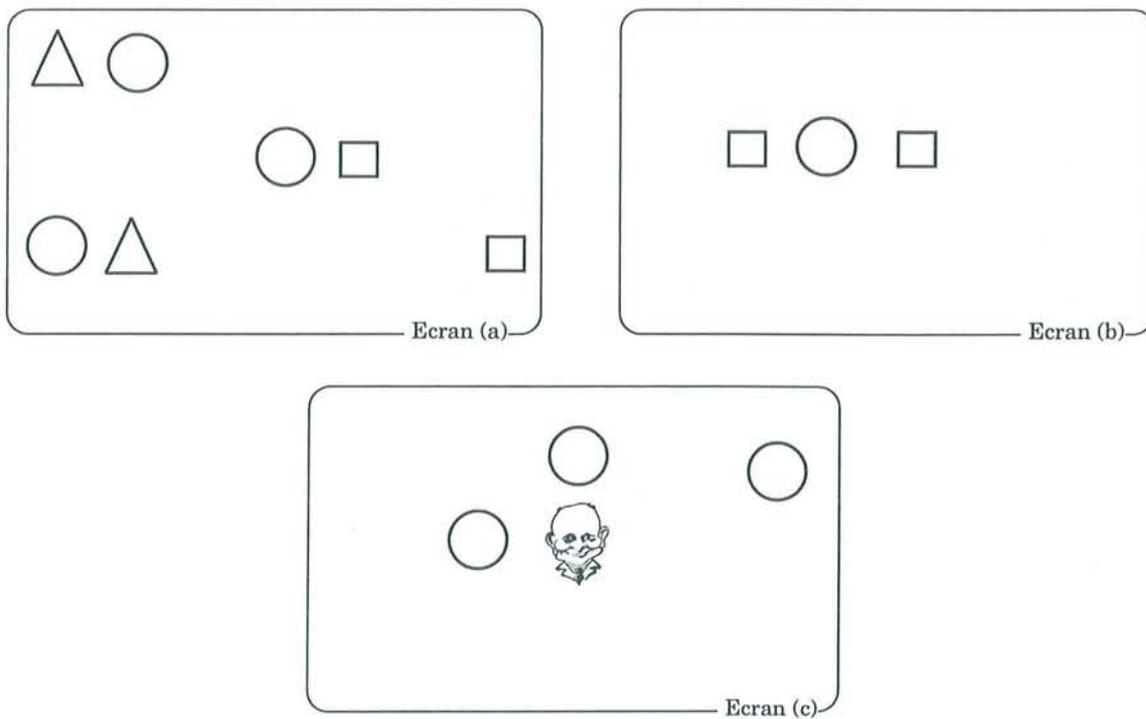


FIG. 4.34 – Détermination du champ “prédiction des alternatives”.

rait manipuler des figures géométriques du type cercle, carré, triangle ou rectangle mais également des cliparts²⁸. Dans ce cas, on manipule deux types d’objets qui n’ont pas le même hyperonyme²⁹. Considérons à présent la requête “*détruis le cercle qui se trouve au-dessus du bonhomme*” qui serait formulée à la figure 4.34.(c). Dans ce cas, cette requête s’instancie dans un cadre de référence qui contient des objets “cercle” et “clipart”, donc dans un ensemble qui contient des objets du type $\{C\} \cup \{S\}$. Cet ensemble est différent de l’ensemble des types $\{C\} \cup \neg\{C\} \cup \{S\} \cup \neg\{S\}$ que nous avons proposé pour la requête “*détruis le carré qui se trouve à droite du cercle*”. Cette remarque nous incitera donc à ne pas aller au delà de la prédiction $\neq\{C\}$ et $\neq\{S\}$ pour le contenu du nouveau cadre de référence au moment de son instanciation sur une zone de l’écran.

d) Le champ “prédiction d’évolution”

Cet item a un rôle à jouer dans l’élaboration du nouveau cadre de référence en regard du précédent cadre de référence. En effet, certains indices linguistiques peuvent nous renseigner sur une évolution probable du cadre de référence. Plus précisément, les opérations coûteuses d’un point de vue cognitif (agrandissement et rupture) sont souvent accompagnées soit d’un temps de latence, soit d’un véritable marqueur linguistique. Ce

²⁸. Clipart : image. Il s’agit bien souvent de petits dessins vectoriels que l’on peut agrandir et rétrécir à sa guise, sans perte de qualité.

²⁹. L’hyperonyme dépend fortement de la tâche considérée. Nous supposons dans ce exemple que les hyperonymes de “cercle” et “clipart” sont respectivement “figure géométrique” et “dessin” ou “figure vectorielle”.

fait apparaît dans l'extrait du corpus de Mignot [1993] (p. 204) lorsque l'utilisateur utilise des syntagmes du type “*eah...*”, “*à présent*”, etc, qui sont parfois précédés d'un certain temps de latence (marqué [...] dans le corpus).

L'ensemble de ces marqueurs possède un caractère prédictif qui nous renseigne sur le fait que l'utilisateur effectue probablement une réinitialisation dans la conception qu'il a de la scène affichée, cette réinitialisation s'accompagne généralement d'un temps de réflexion qui apparaît dans le dialogue. Ce temps de réflexion permet alors de concevoir la scène sous un angle différent, typiquement, un *agrandissement* ou une *rupture*, les opérations comme la *conservation* ou la *focalisation* requérant quant à elles des temps de réflexion bien moindres car elles se font dans la continuité du raisonnement.

Il convient cependant d'être prudent sur le caractère prédictif des temps de latence et des marqueurs “*maintenant*” et “*à présent*”. Ils peuvent en effet très bien coïncider avec une simple hésitation de l'utilisateur, ce dernier ne sachant précisément que faire ou comment le réaliser à un instant donné. Tout au plus pouvons nous présupposer que ses hésitations seront plus longues lorsqu'il conçoit effectivement la scène affichée sous un autre angle (cette remarque nécessiterait bien entendu une validation psycholinguistique).

Par ailleurs, les déterminants “*celui de*” ou “*celle de*” ou “*ceux de*” recherchent nécessairement leur(s) référent(s) dans le précédent focus ou le précédent cadre de référence. Ainsi que nous avons pu le montrer au précédent paragraphe concernant le champ “prédiction d'évolution”, ce fait provient de la nature de ces marqueurs qui cherchent à isoler un ou plusieurs référents du discours parmi les référents précédemment isolés.

Nous proposons donc la table de correspondance suivante (qu'il conviendra d'étendre par la suite) entre certains indices linguistiques et l'évolution probable du cadre de référence. Ce tableau confère à notre modèle un caractère prédictif.

<i>Indice linguistique</i>	<i>Prédiction de l'évolution du cadre de référence</i>
[...] “ <i>à présent</i> ” “ <i>maintenant</i> ” “ <i>eah...</i> ”	Agrandissement ou rupture Agrandissement ou rupture Agrandissement ou rupture Agrandissement ou rupture
“ <i>celui de</i> ”, “ <i>celle de</i> ”, “ <i>ceux de</i> ”	Focalisation dans le précédent focus ou conservation dans le précédent cadre de référence.

Nous avons pu illustrer à la figure 4.33 pourquoi les références en “*celui de*”, “*celle de*” et “*ceux de*” pouvaient s'effectuer tantôt au sein du précédent focus ou tantôt au sein du précédent cadre de référence.

e) Ellipses et ambiguïtés anaphoriques

Pour les requêtes auxquelles nous nous intéressons qui sont du type: <marqueur>* <verbe> <déterminant cible>* <cible> <séparateur>* <préposition spatiale> <déterminant site>* <site>* (donc sans adjectifs), nous n'avons pas recensé

de tournures elliptiques comme celle qui apparaît par exemple dans la phrase (156). Cependant, dans un souci de généralité et, en vue d'extensions futures de ce modèle, nous mentionnons à ce niveau, l'ajout possible d'un module de traitement des ellipses. La résolution des ellipses s'effectue traditionnellement par le biais d'un historique du dialogue qui fait lui même partie du module de gestion du dialogue (le contenu d'un module de gestion du dialogue sera détaillé au chapitre 5).

Agrandis les carrés de droite. (155)

Détruis le plus petit. (156)

Si le problème de l'ellipse ne nous préoccupe pas ici, en revanche, certaines ambiguïtés anaphoriques peuvent subsister ainsi que cela apparaît dans la requête (158) où il faut que l'analyseur syntaxico-sémantique puisse déterminer que l'anaphore porte sur un carré et non sur un triangle par exemple :

Agrandis le carré de droite. (157)

Détruis celui de gauche. (158)

Là encore, le recours à l'historique du dialogue s'avère nécessaire afin que l'analyseur puisse renvoyer le 11-uplet suivant associé à la requête (158) :

- (1) Marqueur : ×,
- (2) Verbe : “*détruire*”,
- (3) Déterminant de la cible : “*celui*”,
 - genre : masculin,
 - nombre : singulier,
 - type : démonstratif,
- (4) Catégorie de la cible : “*carré*”,
- (5) Séparateur : ×,
- (6) Déterminant du site : ×,
 - genre : ×,
 - nombre : ×,
 - type : ×,
- (7) Catégorie du site : ×,
- (8) Préposition spatiale : “*de gauche*”,
 - type de préposition spatiale : simple,
- (9) Type de localisation : absolue,
- (10) Prédiction des alternatives : “*carré*”,
- (11) Prédiction d'évolution : focalisation ou conservation.

Pour un descriptif détaillé d'une solution possible à cette résolution anaphorique, on pourra se reporter aux travaux de Gaiffe [1992].

f) Conclusion

Cette section nous a permis de définir le rôle et le calcul des champs “prédiction des alternatives”, “prédiction d'évolution”. Par ailleurs, nous avons également pu décrire dans cette partie la construction des champs “catégorie du(des) site(s)” et “catégorie de la(les) cible(s)”.

4.5.4 La perception

Nous avons souligné au chapitre précédent l'importance que pouvait jouer la perception visuelle pour le problème spécifique des références spatiales plurielles car, ce type de référence, encore plus que les références au singulier, impose une forte structuration de la scène affichée, pour retrouver les bons référents. Ce point justifie que nous nous intéressions à présent à l'activité de perception humaine qui est une activité de nature psychologique. L'objectif de notre étude est de dégager certains éléments que nous transposerons dans le domaine du dialogue homme-machine.

Delay et Pichot [1984] reprennent la définition proposée par Piéron [1961] pour définir la perception comme étant «*la prise de connaissance sensorielle d'événements extérieurs qui ont donné naissance à des sensations plus ou moins nombreuses et complexes. La perception fournit un percept*³⁰.» Un percept n'est pas seulement la somme des sensations mais bien plus, il est caractérisé par son intensité, sa netteté, sa clarté et sa richesse en détails [Denis, 1989].

a) Les trois composantes de la perception

Delay et Pichot [1984] distinguent schématiquement dans chaque perception trois composantes, étroitement associées :

a.1) *Le processus récepteur*

Ce processus permet de passer de l'ensemble des sensations au percept. Ce passage s'opère en distinguant dans le champ perceptif les différents éléments constituant le percept sur la base d'une discrimination fondée sur la forme, la couleur et la taille par rapport au fond.

a.2) *Le processus symbolique*

À chaque champ structuré est associé un concept. Ainsi, si l'on regarde une pomme qui est placée sur une table, nos processus récepteurs nous permettent de distinguer un champ structuré, comportant un objet sphérique, d'une certaine taille et d'une certaine couleur qui se détache du fond constitué par la table. Au même moment, nous percevons que cet objet est une pomme. L'opération réalisée est une appréhension immédiate d'une

30. Percept : objet de la perception, sans référence à une chose en soi (opposé à concept).

représentation abstraite qui contient un ensemble de propriétés.

a.3) *Le processus affectif*

Ce processus illustre le fait que toute perception comporte des aspects affectifs. En particulier, une expérience perceptive peut être agréable, désagréable ou indifférente.

b) Les lois de la perception

Les psychologues gestaltistes ont énoncé un certain nombre de lois perceptives pour rendre compte du fait que certaines organisations perceptives ont plus de chances de se produire que d'autres. Ces lois sont toutes gouvernées par le principe selon lequel «*le tout est différent de la somme des parties*». Elles dépendent d'une part de l'expérience de l'individu et d'autre part du champ perceptif. Voici les principales lois perceptives :

1. *Le principe de structuration.* Les éléments perceptifs isolés ont une tendance spontanée à la structuration. Cette structuration se fait prioritairement suivant le principe de proximité et le principe de similitude :

- *Le principe de proximité.* Les choses proches sont regroupées ensemble.
- *Le principe de similitude.* Les choses qui ont l'air «*semblables*» sont regroupées. Ainsi, à la figure 4.35 tirée de [Bruce et Green, 1993], nous percevons des colonnes, même si l'information de proximité suggère plutôt des rangées. Ceci illustre le fait que la similitude peut prendre le pas sur l'information de proximité. Le fait de savoir jusqu'où des items doivent être semblables pour être regroupés ensemble reste une question empirique [Bruce et Green, 1993].

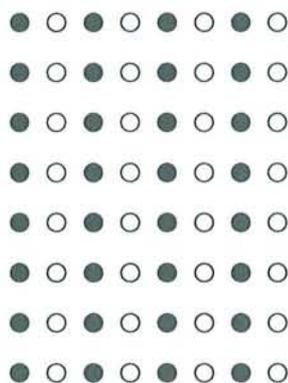


FIG. 4.35 – Image perçue comme étant composée de colonnes, la similitude des brillances des points l'emporte sur la proximité.

2. *Le principe de discrimination entre le fond et la forme.* Les formes ou *Gestalten* tendent à se détacher comme ensembles limités, structurés, ayant une unité subjective constituant une figure qui se détache du fond non structuré [Delay et Pichot, 1984].

3. *Le principe d'association entre la forme et la signification.* Ce que l'on voit est déterminé en partie par ce que l'on s'attend à voir, étant donné son vécu personnel.
4. *Le principe de constance.* Les bonnes formes tendent à garder leurs caractéristiques propres malgré les modifications de la présentation. Pour illustrer ce principe, il suffit de considérer par exemple le charbon qui sera perçu comme noir, même si nous le voyons en plein soleil. De la même manière, la neige sera perçue comme blanche et ce, même au crépuscule.
5. *Le principe d'actualisation des formes.* Toutes les formes se "développent" ou s'actualisent à partir de préformes jusqu'à des formes qui sont prégnantes et bien différenciées.
6. *Le principe de la prégnance.* Ce principe subsume en quelque sorte les autres principes. Il détermine la facilité avec laquelle une forme est perçue comme une figure se détachant par rapport au fond. On peut résumer ce principe comme suit : de plusieurs organisations géométriques possibles émergera celle qui possède la forme la meilleure, la plus simple et la plus stable.

Par ailleurs, les perceptions complexes comme celles de la profondeur et du relief mettent en œuvre huit facteurs différents et d'importance inégale :

1. La perception kinesthésique du *degré de contraction du cristallin* au cours de la fixation de l'objet lors de l'accommodation.
2. La perception kinesthésique provenant des muscles moteurs de l'œil au cours de la *convergence* des deux axes oculaires sur l'objet observé.
3. La perception visuelle de la *disparité des images rétiniennes* d'un même objet pour les deux yeux. Cette disparité est d'autant plus grande que l'objet est rapproché.
4. La perception visuelle de la modification que subit l'image rétinienne lors de mouvements de rotation de la tête. La variation de position de l'image est d'autant plus grande que l'objet est proche.
5. La perception des *effets de recouvrement* : un objet placé devant un autre le cache partiellement.
6. La perception visuelle de la *perspective aérienne* : les objets éloignés ont des contours plus flous que les objets rapprochés.
7. La perception de *l'effet de perspective* qui traduit la structure du champ dans lequel s'intègre l'objet.
8. La *connaissance de la taille approximative* des objets résultant de nos expériences antérieures.

c) Application à un système de dialogue homme-machine

L'étude qui précède nous fournit une structuration *a priori* de la scène affichée, c'est-à-dire qu'il s'agit d'une structuration qui s'opère de manière spontanée. On peut reprocher à ce type de structuration qu'elle opère pour ainsi dire "sans âme", en effet, l'utilisateur peut très bien projeter un point de vue sur la scène affichée. Pour s'en convaincre, il suffit de reconsidérer la figure 4.35 (p. 174) pour une requête du type "*détruis les cercles d'en haut*" qui induit alors une forte opposition spatiale entre les huit cercles qui se trouvent affichés tout en haut de la figure et les autres. Le groupe prépositionnel "*d'en haut*" bouleverse alors la structuration spontanée de la scène perçue.

Dans le cadre du dialogue homme-machine, l'étude précédente et la réflexion de [Bellalem, 1995] nous conduisent à distinguer trois capacités, pour un humain de structurer et d'organiser la scène qu'il a sous les yeux :

c.1) La capacité à percevoir des groupes d'objets

La formation de groupes homogènes perceptifs répond à deux caractéristiques. Pour rassembler deux ou plusieurs objets, il faut qu'ils présentent une *proximité spatiale* et une *similitude sur au moins une propriété*. La similitude peut être purement perceptuelle suivant le type des objets, leur couleur, leur forme, leur taille, leur brillance. La similitude peut également être spatiale (exemple (a) ci-après). Enfin, la similitude peut être fonctionnelle (exemple (b) ci-après). Voici respectivement un exemple de constitution de groupe homogène perceptif sur la base d'une propriété spatiale et d'une propriété de similitude fonctionnelle :

(a) *deux livres superposés*: la proximité spatiale est liée à la superposition des livres et la similitude concerne le type des objets (*livre*).

(b) *un téléphone et un ordinateur posés sur un bureau*: la proximité spatiale est également manifeste dans ce cas du fait que le téléphone et l'ordinateur sont posés sur un même bureau. Par ailleurs, la propriété fonctionnelle qui unit ces trois objets provient du fait que deux d'entre eux entretiennent la même relation avec un même troisième (la relation de *superposition*)³¹.

On pourra remarquer que les groupes homogènes perceptifs que nous avons proposés au chapitre précédent se construisent en fait sur la base d'une proximité spatiale et d'une ou plusieurs propriétés non fonctionnelles. En particulier, les groupes de Briffault (3.32.(b) à la page 118) sont construits à l'aide d'une proximité spatiale et d'une propriété de type commune au groupe. Enfin, pour les groupes homogènes perceptifs que nous avons proposés dans le précédent chapitre, tandis que les groupes de type *droite/gauche* adjoignent à ces deux propriétés une proximité horizontale, les groupes de type *haut/bas* y adjoignent une propriété de proximité verticale.

c.2) La capacité à percevoir différents niveaux de détail ou granularité

31. Les exemples (a) et (b) sont tirés de [Bellalem, 1995].

Cette capacité exprime le fait que l'utilisateur peut percevoir la scène à différents degrés de précision. Ainsi, pour la figure 4.36, il peut très bien concevoir cette dernière à un niveau de détail élevé en ne structurant sa scène que sous la forme d'icônes et de fenêtres. Au contraire, il peut vouloir agir sur ces dernières en considérant un niveau de détail plus fin pour s'intéresser aux boutons ou aux ascenseurs qui sont contenus dans les fenêtres par exemple.

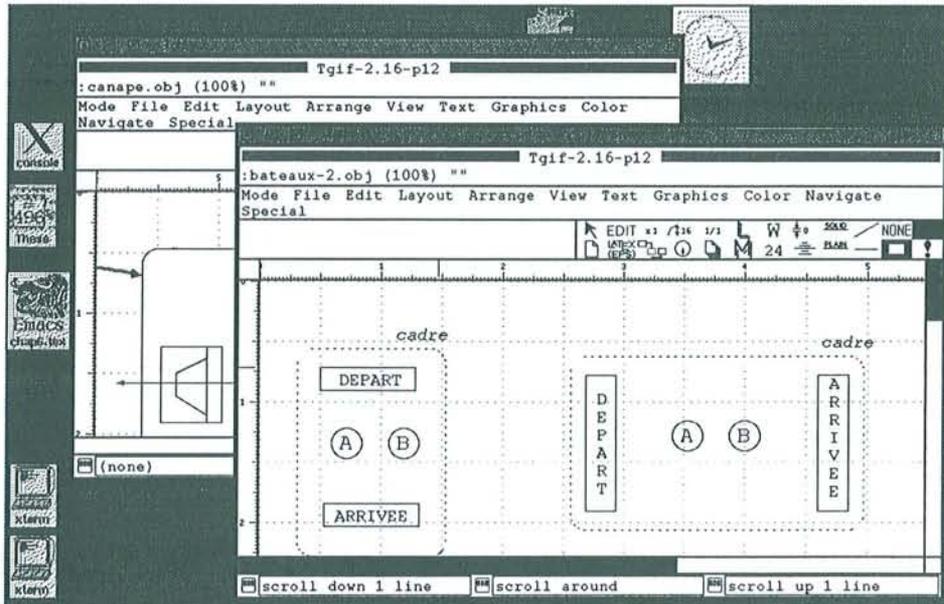


FIG. 4.36 – Un système de multifenêtrage.

La capacité à percevoir différents niveaux de détail ou de granularité s'appuie directement sur la capacité à produire un filtre visuel.

c.3) La capacité à produire un filtre visuel

Cette faculté permet à l'utilisateur d'avoir un point de vue sur la scène en y recherchant uniquement des éléments d'un type donné et en identifiant les relations spatiales qui lient les différents candidats. La capacité à produire un filtre visuel permet alors de concevoir la scène à un certain niveau de détail.

Ce principe nous sera très utile pour ne prendre en compte que les objets qui sont effectivement susceptibles d'être considérés par la référence spatiale. Ainsi à la requête (159), si le cadre de référence doit s'instancier au départ sur une portion d'écran qui contient des "carrés" et des "carrés", un filtrage du cadre de référence pour qu'il ne contienne plus que des carrés nous permettra d'isoler effectivement le groupe homogène perceptif, contenant des carrés, qui se trouve le plus à droite.

Détruis les carrés de droite.

(159)

4.5.5 L'application

On peut définir l'application comme un ensemble de compétences logicielles disponibles, en vue de réaliser des tâches. Ces compétences vont permettre à l'utilisateur d'atteindre un but en passant d'un état initial à un état final. Cette évolution s'opère par le biais d'actions successives sur l'état courant de la tâche.

C'est à ce niveau en particulier que sont définies les conditions d'application des actions sur les objets ainsi que leurs effets.

a) La représentation de l'application

L'utilisateur va agir sur son application par le biais d'une scène d'objets qui est visualisée sur son écran. Si cette scène ne matérialise bien souvent qu'une partie de l'application³², on pourra cependant remarquer que l'essentiel des références spatiales portera sur cette scène. Nous insistons sur le mot "l'essentiel" car on peut très bien imaginer certaines requêtes de l'utilisateur qui feraient intervenir des références spatiales qui ne font pas appel à la scène affichée ou pas entièrement. Il suffit de prendre l'exemple de "*Imprime ce fichier sur l'imprimante lj1*" qui se fonde sur une référence spatiale mais qui ne fait pas intervenir de relation de superposition à l'écran et ce, malgré l'usage de la préposition "*sur*".

Par ailleurs, le type de représentation qui est utilisé pour visualiser la scène est très important. On distingue classiquement les représentations 2D, $2D\frac{1}{2}$ (qui mettent en œuvre la notion d'épaisseur ce qui autorise le chevauchement des objets (figure 4.36)) et enfin les représentations 3D qui permettent d'obtenir des scènes très réalistes en s'appuyant sur des notions de relief, de nuance des couleurs, de luminosité, etc.

Il semble important de remarquer que le panel des expressions référentielles dont l'utilisateur fera usage dépend nécessairement du type de représentation utilisé. Les représentations 2D autorisent de manière prototypique des expressions référentielles composées des prépositions "*à droite de*", "*à gauche de*", "*au-dessus de*", "*en dessous de*" et "*à côté de*", les représentations $2D\frac{1}{2}$ autorisent en plus les prépositions "*devant*", "*derrière*", "*sur*", "*sous*", enfin les représentations 3D autorisent pratiquement toute latitude possible à l'utilisateur quant au choix des prépositions spatiales.

b) Les contraintes fonctionnelles

C'est à ce niveau que doivent être décrites l'applicabilité ou l'inapplicabilité de tel ou tel prédicat sur tel ou tel objet de l'application. Il est par exemple tout à fait possible qu'un objet donné ne soit pas déplaçable dans l'application considérée. Une requête du type "*déplace le rectangle de droite*" pour peu qu'elle soit ambiguë *a priori* pourrait être désambiguïsée si l'un des objets sur lesquels porte la référence possède le caractère d'être non déplaçable par exemple.

32. L'application peut faire intervenir notamment des liens fonctionnels entre les différents objets, ces liens n'apparaissant pas nécessairement à l'écran.

Nous rappelons ici les contraintes fonctionnelles que nous nous proposons de considérer :

- les relations de contact,
- les relations d'inclusion :
 - inclusion totale,
 - inclusion partielle :
 - inclusion classique,
 - inclusion partie-tout de type composant-assemblage.

4.6 Principe de fonctionnement de la notion de cadre de référence

La théorie des cadres de référence vise à construire pour chaque requête de l'utilisateur le nouveau cadre de référence associé qui permettra de retrouver la ou les cibles recherchée(s) à l'intérieur du focus. La construction effective du cadre de référence en situation de dialogue s'effectue en plusieurs étapes (revoir la figure 4.15) :

1. Tout d'abord, la requête de l'utilisateur est analysée par le module syntaxico-sémantique qui élabore le 11-uplet. Ce 11-uplet fournit deux champs très importants qui sont le champ "prédiction des alternatives" et "prédiction d'évolution". Le premier indique la nature des objets qui doivent être contenus au départ dans le cadre de référence lors de son instanciation effective sur la scène affichée. Le second fournit une prédiction quant au type d'évolution qui sera probablement réalisé par rapport au précédent cadre de référence (ces deux prédictions ne sont pas toujours vérifiées). Par ailleurs, des *méta-règles* sur lesquelles nous travaillons actuellement permettent d'ajuster éventuellement la prédiction d'évolution. Par exemple, une prédiction de *rupture* qui suivrait une *rupture* ayant déjà eu lieu au sein du précédent cadre de référence sera modérée. D'autre part, une *focalisation* sera bien souvent suivie d'un *agrandissement* ou d'une *rupture*. Ce fait nous incitera à renforcer ces deux possibilités si l'une d'elles apparaît dans la champ "prédiction d'évolution" et que l'ancien cadre de référence provient effectivement d'une *focalisation*.

2. La nature de la(des) cible(s), la préposition spatiale mise en jeu dans la requête de l'utilisateur ainsi que la nature de l'application peuvent induire des contraintes fonctionnelles entre le(s) site(s) et la(les) cible(s). L'application nous renseigne également sur l'allure de l'image mentale du nouveau cadre de référence, sera-t-il en 2D, en $2D\frac{1}{2}$ ou en 3D.

3. Le lexique fournit l'allure du focus de recherche associé à la préposition spatiale et l'(les) orientation(s) pertinente(s) qui fera(feront) partie du cadre de référence.

4. La scène est filtrée pour ne retenir que les objets qui sont du même type que la(les) cible(s) et le(s) site(s) éventuel(s)³³. Le nouveau cadre de référence qui est en

33. Il est nécessaire de parler de "site(s) éventuel(s)" car certaines requêtes ne comportent pas de site explicite comme par exemple : "détruis le carré de droite".

phase de construction se réduit alors aux objets filtrés. En cas de référence plurielle, la scène est également structurée perceptivement selon la préposition spatiale qui est mise en jeu.

À l'issue des différentes contributions que sont l'analyse syntaxico-sémantique, les apports de l'application, les apports provenant de la perception visuelle mais aussi les connaissances provenant du lexique et de l'historique du dialogue, on obtient finalement un nouveau cadre de référence qu'il convient d'instancier sur une partie de l'écran ou sur la totalité de ce dernier. Voyons à présent le détail de la recherche de la(les) cible(s) et du(des) site(s) :

1. Suivant la valeur du champ "prédiction d'évolution" qui peut être de quatre types : "focalisation dans le focus", "conservation dans le précédent cadre de référence", "agrandissement" et "rupture", l'instanciation du nouveau cadre de référence sera tentée pour le premier type d'évolution au sein du précédent focus et pour les trois autres au sein du précédent cadre de référence. Ces instanciations devront bien entendu valider le champ "prédiction des alternatives". Le nouveau cadre de référence occupera donc au pire la totalité de l'écran. On se rappellera à ce propos la phrase "*détruis le carré de droite*" que nous avons décrite dans ce chapitre si elle était énoncée sur un écran ne contenant que des carrés. Le déterminant "*le*" forcerait le champ "prédiction des alternatives" à la valeur $\{C\} \cup \neg\{C\}$ c'est-à-dire "*carré*" \cup "*non carré*". L'instanciation du nouveau cadre de référence serait alors impossible dans l'absolu et ce, même sur la totalité de l'écran car ce dernier ne contient pas d'objets du type "*carré*" et "*non carré*". Nous avons pu faire remarquer que l'écran ne contenait pas que des carrés. Au contraire, il serait par exemple possible de parler du "fond de l'écran" ce qui rend alors légitime la requête "*détruis le carré de droite*". Dès lors, un cadre de référence dont le champ "prédiction des alternatives" est égal à $\{C\} \cup \neg\{C\}$ qui devrait être instancié sur une scène qui ne contient que des objets d'un type $\{C\}$ donné, le sera finalement sur la totalité de l'écran.

2. Une fois le cadre de référence instancié sur une partie de l'écran ou sur la totalité de ce dernier, on filtre les éléments inclus pour ne retenir que ceux qui sont du type de la(des) cible(s) et du(des) site(s), le nouveau cadre de référence se restreint ensuite à ces objets filtrés. On pourra également filtrer le cadre de référence en ne retenant que la(les) cible(s) sur laquelle le prédicat est effectivement applicable. Il ne reste plus alors qu'à étendre le(s) focus jusqu'à retrouver la(les) cible(s) recherchée(s) (le cas où plusieurs focus sont mis en jeu apparaît lorsque le site est ambigu dans l'absolu, voir à ce titre la figure 4.19.(b) (p. 153) pour la requête "*détruis le cercle qui se trouve à droite du triangle*"). L'étendue du focus de recherche ne sort bien entendu pas du cadre de référence associé.

3. En cas d'échec, on reprend la recherche dans un(des) cadre(s) de référence agrandi(s) de la distance d (qui varie d'un site à l'autre si plusieurs sites sont considérés simultanément). Le(s) cadre(s) de référence est(sont) ainsi agrandi(s) jusqu'à ce qu'il(s) recouvre(nt) la totalité de l'écran. À l'issue de l'agrandissement du(des) cadre(s) de référence prototypique(s), si la(les) cible(s) n'est(ne sont) pas trouvée(s),

il faut reprendre au point 2 avec un nouveau cadre de référence. Ce nouveau cadre de référence sera alors dans un premier temps instancié dans le précédent cadre de référence. Si cette nouvelle recherche échoue, l'algorithme (ses grandes lignes se trouvent à la figure 4.37) reprendra avec le cadre de référence qui l'a précédé (on remonte alors d'un certain nombre de cadres de référence en arrière. Ce nombre est prédéfini et dépend de l'évolutivité de la tâche, en particulier, si cette dernière évolue rapidement les retours en arrière seront limités). Enfin, en cas de nouvel échec, le nouveau cadre de référence sera instancié sur la totalité de l'écran.

4. Une fois la(les) cible(s) localisée(s), il ne reste plus alors plus qu'à déclencher le prédicat sur la(les) cibles obtenues. Si cette thèse porte sur l'étude de la référence aux objets et non sur la référence aux actions, on pourra cependant remarquer comme cela apparaîtra dans l'exemple du paragraphe 4.7 que le prédicat peut avoir une influence sur le focus après l'instanciation du cadre de référence. Il suffit en effet de prendre le requête (160) où, à l'issue de l'agrandissement des carrés, il paraît légitime que le focus englobe les carrés une fois agrandis et ne se limite pas à l'étendue d'espace qu'occupaient les carrés avant l'agrandissement (il faudra bien entendu agrandir également le cadre de référence associé qui englobe nécessairement le focus). Ce fait explique que nous nous soyons cantonné à l'étude de phrases contenant des prédicats dont l'influence est relativement neutre sur le focus.

Agrandis les carrés de gauche. (160)

Finalement, suivant la portion de l'écran qu'occupe effectivement le nouveau cadre de référence par rapport au précédent, on sait alors quel type d'évolution a été réalisé : une conservation, une focalisation, un agrandissement ou une rupture.

On peut donc distinguer quatre étapes au sein du fonctionnement des cadres de référence :

1. l'étape de construction de la prédiction portant sur les alternatives ainsi que l'élaboration de la prédiction d'évolution du cadre de référence,
2. l'étape de construction effective du cadre de référence et son instanciation sur une partie de l'écran ou la totalité de ce dernier. Une étape de filtrage suivie d'une restriction du cadre de référence aux objets filtrés puis une éventuelle structuration de la scène affichée. La recherche de la(les) cible(s) n'est pas nécessairement directe du fait d'éventuelles ambiguïtés portant sur le site,
3. la recherche effective du(des) site(s) et de la(des) cible(s),
4. l'étape qui permet de détecter l'évolution du cadre de référence, d'accomplir l'(les) action(s) associée(s) au prédicat et enfin de redimensionner éventuellement le focus et le cadre de référence englobant associé.

On trouvera à la figure 4.37 l'algorithme qui permet de réaliser ces différentes étapes.

- (1) **Analyse syntaxico-sémantique + méta-règles d'évolution du cadre de référence**
 ➔ construction du 11-uplet
- (2) **Application + nature site(s) + nature cible(s) + nature préposition spatiale**
 ➔ contraintes fonctionnelles cible/site
- Application**
 ➔ type (2D ou 2D1/2 ou 3D) de l'image mentale du cadre de référence
- Lexique**
 ➔ focus + type d'extension
 ➔ cadre prototypique englobant le(s) site(s) (excepté pour le cas de la localisation absolue qui ne fait pas appel à un site explicite)
- (3.A) Si (prédiction d'évolution='focalisation') ou (prédiction d'évolution='conservation') et (précédent focus existe) alors
 alors
 demander à l'utilisateur pour désambigüiser
 suivant le cas : Nouveau cadre de référence=Précédent cadre de référence
 ou Nouveau cadre de référence=Précédent focus
 sinon Nouveau cadre de référence = Précédent cadre de référence
- (3.B) Si (prédiction des alternatives non vérifiée) aller en (3.E)
 sinon
- Perception**
 ➔ filtrage des objets possédant le type de la(les) cibles et du(des) site(s) (s'il(s) existe(nt)) et restriction du cadre de référence autour des ces objets filtrés
- Perception et plusieurs cibles**
 ➔ structuration perceptive de la scène selon la préposition spatiale
- Prédicat + application**
 ➔ filtrage éventuel sur la(les) cible(s)
- (3.C) Agrandissement du focus jusqu'à isoler une cible ou un groupe homogène perceptif de cibles⁽¹⁾
- (3.D) Si cible(s) trouvée(s)
 renvoyer le(s) résultat(s)
 aller en (4)
 Sinon si (localisation-relative) alors
 agrandir le(s) cadre(s) de référence associé(s) au(x) site(s) (arrêt lorsque le(s) cadre(s) recouvre(nt) tout l'écran)
 aller en (3.C)
- (3.E) Reprendre en (3.B) avec cadre de référence courant = précédent cadre de référence (revenir au maximum d'un certain nombre de cadres de référence en arrière, le tout dernier retour se faisant sur la totalité de l'écran)
- (4) **Post-actions**
 ➔ nature de l'évolution du cadre de référence
 ➔ action(s) associée(s) au prédicat
 ➔ modification éventuelle du focus et du cadre de référence englobant.

(1) Si plusieurs sites, agrandir simultanément les focus des différents cadres de référence.

FIG. 4.37 – Algorithme de calcul des références spatiales.

4.7 Application de la notion de cadre de référence à un système de représentation graphique

4.7.1 Introduction

Nous nous proposons d'étudier l'application du modèle présenté ci-dessus au cas particulier du pilotage d'un système graphique pouvant afficher différentes sortes d'objets en deux dimensions comme au précédent chapitre, à savoir, des cercles, des triangles, des carrés et des rectangles. Les objets étudiés ne possèdent donc pas d'orientation intrinsèque. Par ailleurs, aucune contrainte fonctionnelle particulière ne provient de la tâche.

4.7.2 Exemple détaillé

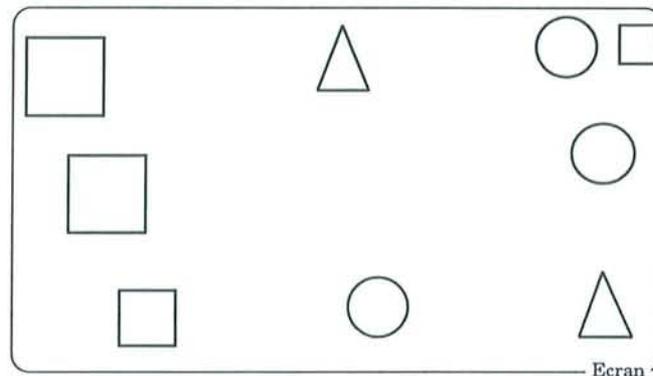


FIG. 4.38 – Une scène d'objets.

L'exemple que nous allons étudier va nous permettre de démontrer que le modèle proposé apporte des solutions au problème du suivi discursif des référents spatiaux en remédiant notamment à certaines absences du modèle proposé au précédent chapitre. Voici la séquence de requêtes à analyser que nous illustrerons pour la figure 4.38 :

1. *Agrandis les carrés de gauche.*
2. *Détruis celui de droite.*
3. *Détruis le carré de droite.*
4. *Détruis celui de gauche.*
5. *Agrandis le cercle qui se trouve au-dessus du triangle.*

Étant donné qu'il n'existe pas de cadre de référence précédent ni d'image mentale associée, nous supposerons que le précédent cadre de référence s'étend sur la totalité de l'écran.

1. *Agrandis les carrés de gauche.*

- (1) **Analyse syntaxico-sémantique** : elle nous renvoie le 11-uplet suivant :
 (1) Marqueur : ×,

- (2) Verbe: “agrandir”,
- (3) Déterminant de la cible: “les”,
 - genre: neutre,
 - nombre: pluriel,
 - type: défini,
- (4) Catégorie de la cible: “carré”,
- (5) Séparateur: ×,
- (6) Déterminant du site: ×,
 - genre: ×,
 - nombre: ×,
 - type: ×,
- (7) Catégorie du site: ×,
- (8) Préposition spatiale: “de gauche”,
 - type de préposition spatiale: simple,
- (9) Type de localisation: absolue,
- (10) Prédiction des alternatives: “carré” \cup ¬“carré”,
- (11) Prédiction d'évolution: *conservation*.

Du fait que nous sommes en présence d'un groupe nominal en “le N”, le champ “prédiction des alternatives” sera instancié par la valeur $\{C\} \cup \neg\{C\}$, en l'occurrence “carré” \cup ¬“carré”.

Il n'y a pas de marqueur linguistique et le déterminant de la cible ne nous apporte pas de prédiction sur l'évolution du cadre de référence: par défaut, le champ (11) est donc instancié comme étant une *conservation*.

• **(2) Application + nature site(s) + nature cible(s) + nature préposition spatiale**: il n'y pas de contraintes cible(s)/site(s) et ce dans la totalité de l'exemple. L'application nous indique également que l'image mentale sera en 2D.

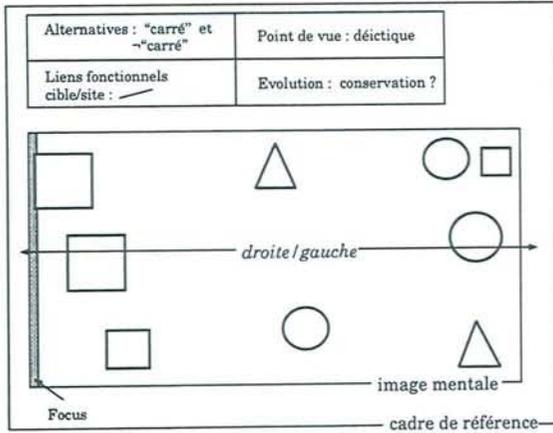
- **Le lexique**: étant donné qu'il n'y a pas de site explicite ici (la préposition “de gauche” jouant le rôle d'un nom de localisation interne à l'intérieur de l'écran), l'étendue du cadre de référence ne sera pas réduite à une zone prototypique. Le lexique nous indique quelle est l'allure du focus et de son extension en se fondant sur la préposition “de gauche”. Le lexique nous indique également que la seule orientation pertinente au sein du cadre de référence est du type *droite* \longleftrightarrow *gauche*. Cet axe d'opposition est orienté à partir de l'orientation déictique qui prévaut dans le type d'application considéré.

• **(3.A)** Le précédent cadre de référence n'existe pas, le champ “prédiction d'évolution” provenant de l'analyse syntaxico-sémantique est du type *conservation*.

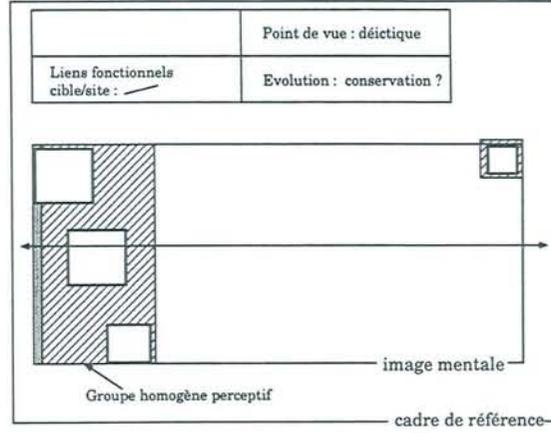
• **(3.B)** le champ alternatives étant égal à la valeur “carré” \cup ¬“carré”, le nouveau cadre de référence sera donc instancié sur la totalité de l'écran (figure 4.39.(1.a)).

- **Perception visuelle**: ses apports sont de deux types. La scène est tout d'abord filtrée pour ne faire apparaître que des carrés, il s'en suit une restriction de l'étendue

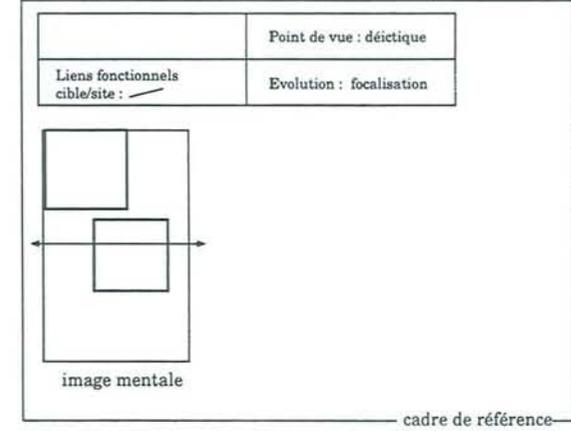
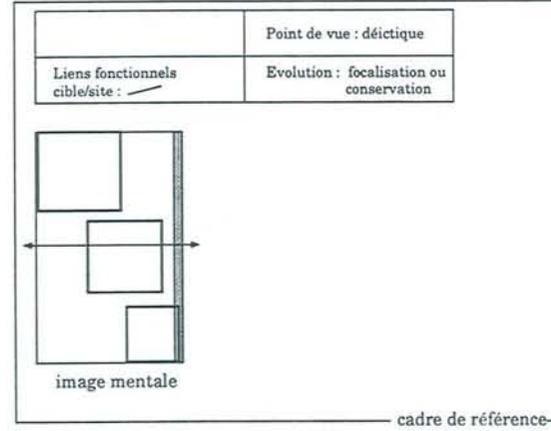
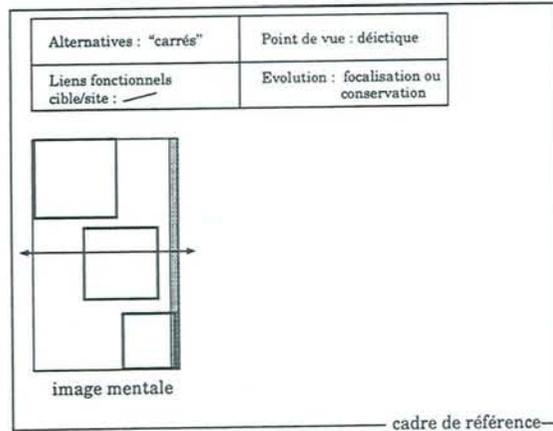
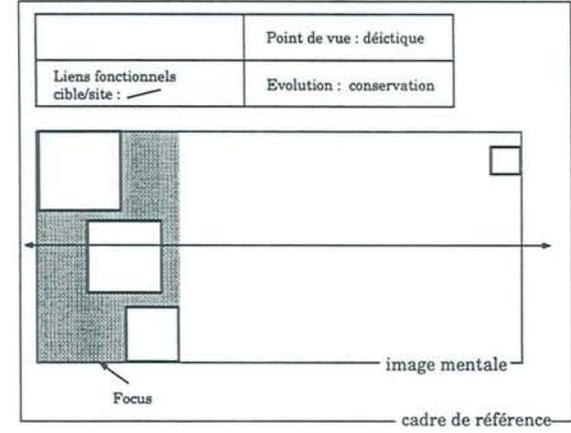
Construction du cadre de référence + focus



Filtrage + restriction + structuration + affinage du focus



Action du prédicat + nouveau focus + nouveau cadre



1. Agrandis les carrés de gauche.
2. Détruis celui de droite.

FIG. 4.39 – Exemple de fonctionnement des cadres de référence.

du nouveau cadre de référence à ces derniers (4.39.(1.b)). Dans un second temps, du fait que nous recherchons un groupe de carrés, il nous faut structurer perceptivement la scène affichée selon des groupes homogènes perceptifs de type *gauche/droite* (figure 4.39.(1.b)).

- (3.C) L'agrandissement progressif du focus nous conduit à isoler immédiatement le groupe perceptif formé des carrés de gauche (figure 4.39.(1.b)). Le groupe de carrés trouvé, il ne reste plus qu'à aller en (4).

- **Post-actions :**

- Nous avons souligné à la précédente section que l'application du prédicat pouvait avoir une influence sur le focus. C'est le cas ici où il faut que le focus englobe l'ensemble des carrés agrandis. Quant au cadre de référence qui doit nécessairement englober le focus, il est nécessaire de l'agrandir afin que le focus y soit strictement inclus.

- Étant donné que l'étendue du nouveau cadre de référence est sensiblement la même que celle du précédent cadre, l'opération d'évolution qui a été effectivement réalisée est une opération de *conservation*. On pourrait sans doute nous rétorquer que le syntagme "*les carrés de gauche*" réalise plutôt une *focalisation* sur les carrés de gauche. Il n'en est rien car l'image mentale du cadre de référence doit contenir l'ensemble des objets pertinents. Ces objets pertinents sont ici en l'occurrence les trois carrés de gauche et celui de droite car les premiers prennent justement pleinement leur signification du fait de l'existence du carré de droite. En effet, si le carré de droite n'existait pas, l'utilisateur aurait tout simplement énoncé "*agrandis les carrés*".

- Les carrés de gauche sont effectivement agrandis d'une certaine taille (ce type d'information qui est attaché au prédicat sort du cadre de cette thèse).

2. Détruis celui de droite.

- (1) **Analyse syntaxico-sémantique :** elle nous renvoie le 11-uplet suivant :

- (1) Marqueur : ×,
- (2) Verbe : "*détruire*",
- (3) Déterminant de la cible : "*celui*",
 - genre : masculin,
 - nombre : singulier,
 - type : démonstratif,
- (4) Catégorie de la cible : "*carré*",
- (5) Séparateur : ×,
- (6) Déterminant du site : ×,
 - genre : ×,
 - nombre : ×,
 - type : ×,
- (7) Catégorie du site : ×,
- (8) Préposition spatiale : "*de droite*",
 - type de préposition spatiale : simple,

- (9) Type de localisation : absolue,
- (10) Prédiction des alternatives : “carré”,
- (11) Prédiction d'évolution : *focalisation* dans le précédent focus ou *conservation* dans le précédent cadre de référence.

Le déterminant “*celui de*” permet d'instancier les champs “prédiction des alternatives” et “prédiction d'évolution”. Le nouveau cadre de référence est alors instanciable soit dans le précédent focus, soit dans le précédent cadre de référence. Il convient dans ce cas d'interroger l'utilisateur pour lui proposer de lever l'ambiguïté. Supposons dans la suite qu'il a choisi le carré qui fait partie du focus.

- **Application + nature site(s) + nature cible(s) + nature préposition spatiale** : leurs apports sont identiques à ceux de la phrase précédente.

- Le **lexique** renvoie l'image mentale du focus de la préposition “*de droite*” ainsi que la nature de son extension. Là encore, il n'y a pas de cadre de référence prototypique associé et la seule orientation pertinente est du type *droite* \longleftrightarrow *gauche*.

- **(3.A)** Nouveau cadre de référence = focus du précédent cadre de référence.

- **(3.B)** - la prédiction est vérifiée (figure 4.39.(2.a)),

- le filtrage et la restriction aux objets filtrés n'apportent rien de plus.

- **(3.C)** l'agrandissement du focus permet de localiser un unique carré (figure 4.39.(2.b)). Le carré étant trouvé, il ne reste plus qu'à aller en **(4)**.

- **(4) Post-actions** :

- Le carré ayant été détruit, le focus associé n'existe plus. Cependant l'étendue du cadre de référence reste inchangée (figure 4.39.(2.c))³⁴.

- Le nouveau site appartient à l'ancien cadre de référence et cadre-réf' \subset cadre-ré³⁵. L'évolution réalisée par rapport au précédent cadre de référence est donc une *focalisation*.

Le traitement des requêtes suivantes sera examiné d'une manière plus rapide en ne nous intéressant qu'aux différences significatives par rapport aux précédentes requêtes.

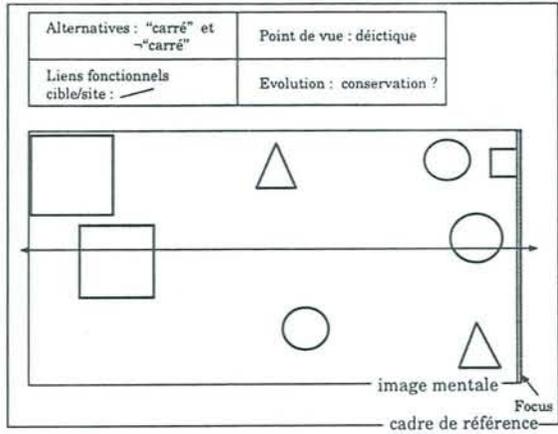
3. Détruis le carré de droite.

Faute d'indice(s) linguistique(s), le champ “prédiction d'évolution” sera instancié par la valeur *conservation*. Le syntagme “*le carré*” nous indique que le champ “prédiction des alternatives” sera instancié par la valeur “*carré*” \cup “*carré*”. Il en résulte en particulier que

34. Nous avons choisi l'option de ne pas modifier l'extension du cadre de référence si le focus, qui s'y trouve inclus, est détruit. En effet, la destruction du focus à la figure 4.40.3.(c) et la restriction qui s'en suivait sur l'étendue du cadre de référence nous empêcherait de détecter qu'un agrandissement a eu lieu entre la figure 4.39.2.(c) et 4.40.3.(c).

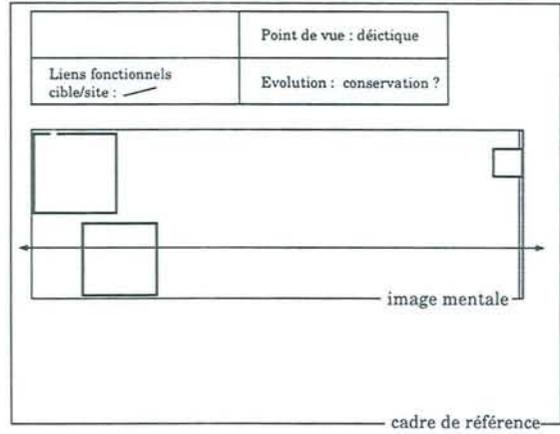
35. *cadre-réf* désigne l'ancien cadre de référence, *cadre-réf'* désigne le nouveau cadre de référence qu'il faut calculer.

Construction du cadre de référence + focus



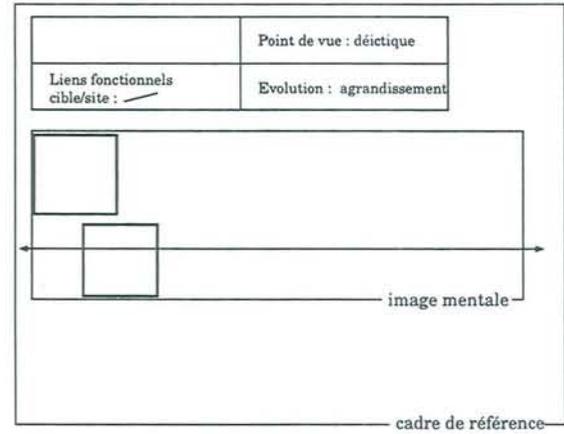
(3.a)

Filtrage + restriction + structuration + affinage du focus



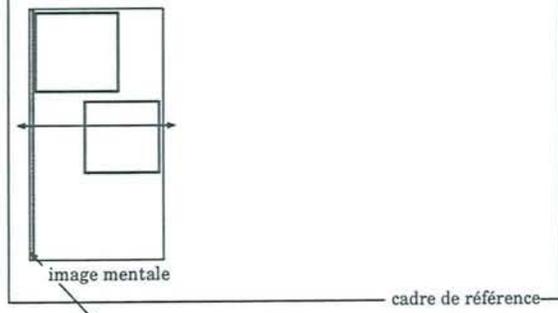
(3.b)

Action du prédicat + nouveau focus + nouveau cadre



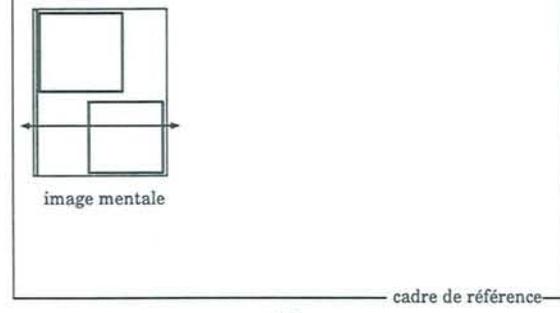
(3.c)

Alternatives : "carrés"	Point de vue : déictique
Liens fonctionnels cible/site : /	Evolution : focalisation ?



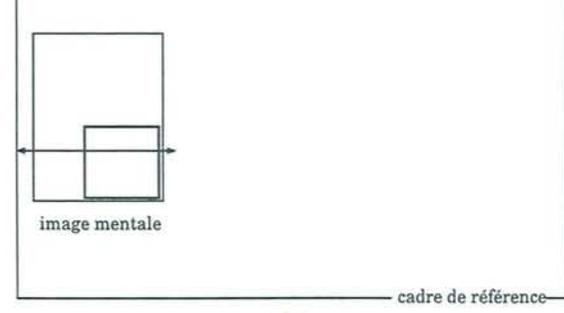
(4.a)

	Point de vue : déictique
Liens fonctionnels cible/site : /	Evolution : focalisation ?



(4.b)

	Point de vue : déictique
Liens fonctionnels cible/site : /	Evolution : focalisation



(4.c)

3. Détruis le carré de droite.

4. Détruis celui de gauche.

FIG. 4.40 – Exemple de fonctionnement des cadres de référence.

le nouveau cadre de référence ne pourra pas plus être instancié dans le précédent cadre de référence que dans celui qui le précède car ces derniers ne contiennent que des carrés.

Le nouveau cadre de référence sera donc instancié sur la totalité de l'écran (figure 4.40.(3.a)). Le filtrage du cadre de référence permet de ne conserver que des carrés et l'étendue du nouveau cadre de référence se restreint à ces derniers. L'agrandissement du focus permet de retrouver le bon carré (figure 4.40.(3.b)).

L'action du prédicat permet de détruire le carré considéré. Enfin, le focus disparaît et le cadre de référence reste inchangé (figure 4.40.(3.c)). Les post-actions aboutissent au fait qu'un *agrandissement* a été réalisé par rapport au précédent cadre de référence car $\text{cadre-réf}' \cap \text{cadre-réf} \neq \emptyset \wedge \text{cadre-réf}' \not\subseteq \text{cadre-réf}$.

4. Détruis celui de gauche.

L'analyse syntaxico-sémantique nous renvoie notamment les deux items suivants :

(10) Prédiction des alternatives: “carré”,

(11) Prédiction d'évolution: *focalisation* dans le précédent focus ou *conservation* dans le précédent cadre de référence.

La focalisation à l'intérieur du focus du dernier cadre de référence est impossible car ce dernier a disparu. La tentative d'instanciation dans celui qui le précède est possible (figure 4.40.(4.a)) et l'instanciation du focus permet de retrouver effectivement le carré référencé par l'utilisateur (figure 4.40.(4.b)).

Les post-actions: l'action du prédicat permet de détruire le carré considéré. Le focus disparaît avec l'objet qui a été détruit. Les post-actions aboutissent au fait qu'une *focalisation* a été réalisée du fait de l'étendue restreinte du nouveau cadre de référence par rapport à l'ancien.

5. Agrandis le cercle qui se trouve au-dessus du triangle.

L'analyse syntaxico-sémantique nous renvoie notamment les deux items suivants :

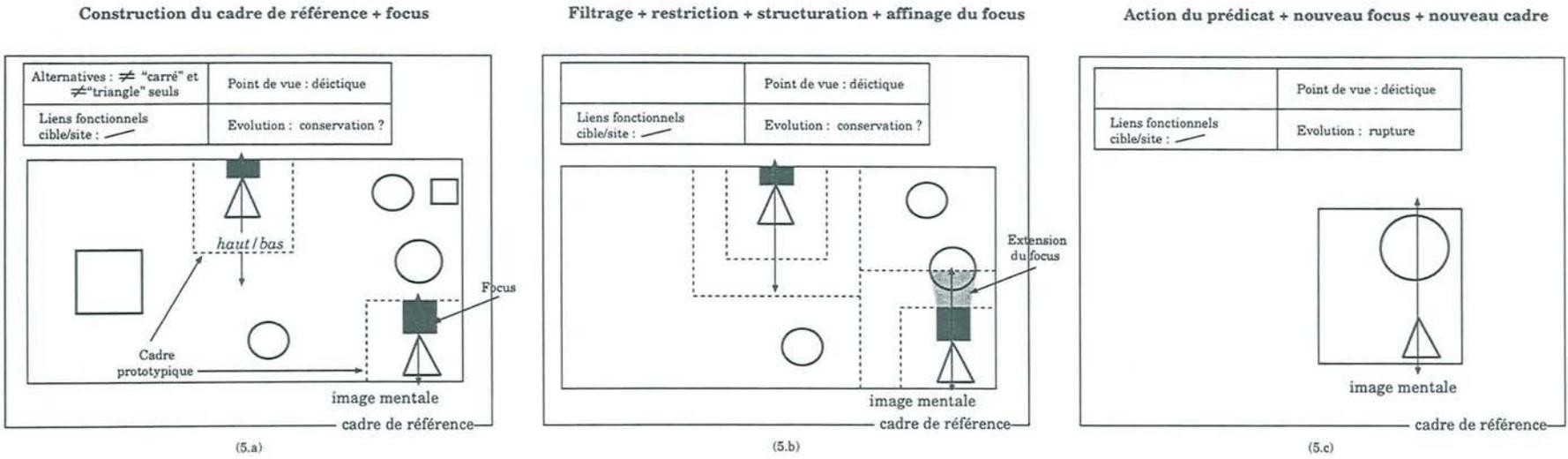
(10) Prédiction des alternatives: $\neq\{C\}$ et $\neq\{S\}$,

(11) Prédiction d'évolution: *conservation*.

L'opération de *conservation* à l'intérieur de l'un des précédents cadres de référence est impossible car ces derniers ne contiennent que des carrés. En revanche (au terme de plusieurs boucles entre (3.B) et (3.E)), une instanciation dans le cadre de référence de départ est possible (c'est-à-dire tout l'écran).

Le lexique nous renvoie la taille du cadre de référence englobant chaque triangle car deux sites sont possibles à la figure 4.41.(5.a). L'agrandissement du focus (point (3.D))

FIG. 4.41 – Exemple de fonctionnement des cadres de référence.



5. Agrandis le cercle qui se trouve au-dessus du triangle.

de l'algorithme) dans chacun des cadres de référence ne permet pas d'isoler un cercle. Il sera donc nécessaire d'agrandir le cadre de référence prototypique associé à chacun des triangles (figure 4.40.(5.b)). Dans ce cas, le focus associé au second triangle permettra effectivement d'isoler le cercle.

Les post-actions : l'action du prédicat permet d'agrandir le cercle considéré ce qui a pour conséquence que le focus et le cadre de référence associé s'agrandissent à leur tour (figure 4.41.(5.c)). Les post-actions permettent donc de déceler une opération de rupture car le nouveau cadre de référence est totalement disjoint du précédent cadre. Il est à noter que, comme nous l'avions indiqué plus haut, une rupture au sein des cadres de référence est difficilement envisageable. C'est effectivement ce qui s'est produit pour la requête "*Agrandis le cercle qui se trouve au-dessus du triangle.*" qui a été prise en compte en deux étapes par les cadres de référence. Une première étape qui nous a conduit à rechercher le triangle et le cercle sur la totalité de l'écran puis finalement une focalisation sur les objets considérés.

4.7.3 Le modèle de la notion de cadre de référence : un bilan

Cette section nous a permis de décrire l'ensemble des connaissances à mettre en œuvre dans un système de dialogue homme-machine à forte composante graphique en vue d'instancier la notion générale de cadre de référence. Le cadre de référence se compose du 5-uplet suivant :

- prédiction des alternatives et prédiction d'évolution du cadre de référence,
- image mentale du focus et de son extension associée,
- point de vue sur le cadre de référence,
- contraintes fonctionnelles éventuelles entre le(les) site(s) et la(les) cible(s),
- évolution effectivement réalisée par rapport au précédent cadre de référence.

Chaque item du 5-uplet est construit à partir de connaissances provenant de l'analyse syntaxico-sémantique, de l'application, de la perception visuelle, mais aussi du lexique et de l'historique du dialogue. La notion de cadre de référence repose sur une représentation duale qui met en œuvre des connaissances relatives à l'imagerie mentale et des connaissances de nature propositionnelle. Les apports de l'approche proposée sont multiples. Tout d'abord, le modèle décrit assure un bon suivi discursif en prenant bien en compte les phénomènes locaux d'interprétation des références spatiales. Ainsi, à un instant donné du discours ne se trouve présente dans le cadre de référence que l'information nécessaire à l'interprétation de la requête de l'utilisateur. Par ailleurs, en cas d'ambiguïté, le modèle change de cadre de référence jusqu'à lever l'ambiguïté. Ces reconsidérations de la scène que peut effectuer le modèle s'apparentent à des changements de point de vue qui permettent de se rapprocher du point de vue adopté par l'utilisateur qui, de son côté ne pense pas avoir émis de requête(s) ambiguë(s).

La notion de cadre de référence parvient alors à calculer les références à des cibles qui sont ambiguës dans l'absolu mais aussi et surtout par rapport à des sites dont la détermination peut également être délicate. Ce dernier point lève la restriction qui pesait sur les algorithmes développés au précédent chapitre où nous avons présumé le caractère

non ambigu du(des) site(s).

Du point de vue de l'implémentation, le modèle proposé nous semble directement implémentable tel quel en vertu de l'algorithme détaillé à la figure 4.37. Ce modèle paraît par ailleurs facilement extensible à d'autres applications à forte composante graphique. Il suffit alors d'adapter le module d'analyse syntaxico-sémantique, les données provenant de l'application ainsi que celles provenant du lexique.

Deux points ont par ailleurs été laissés de côté dans cette section : celui des référents évolutifs et celui de la construction des cadres de référence qui s'opère sur la base de la réunion de plusieurs espaces. Le premier point peut apparaître dans des exemples du type "*iconifie la fenêtre de droite et mets la à gauche*". Dans cet exemple, le genre féminin de l'objet "*fenêtre*" qui a été transformé en "*icône*" est conservé. Il n'en est pas toujours ainsi, le ou les référent(s) deviennent alors particulièrement difficile(s) à retrouver, notamment lorsque leur type est modifié [Gaiffe et Romary, 1995].

Le second point fait partie de la définition de la notion de cadre donnée au paragraphe 4.3.1 : "Cadre : entité qui restreint l'univers à une portion d'espace ou à la connexion de portions d'espace...". La nécessité de recourir à la connexion de portions d'espace n'est pas apparue dans le modèle que nous avons proposé du fait de la nature des requêtes à formaliser et de la tâche étudiée qui nous ont écartés d'exemples aussi complexes que "*Où est Pierre ? Il n'est pas dans son bureau*".

Parmi les extensions à apporter à ce modèle, il convient de mentionner la gestion des conflits entre les différentes orientations (déictique, intrinsèque et extrinsèque) qu'il conviendra de traiter.

4.8 Prémisses pour un modèle de traitement des énoncés de positionnement

4.8.1 Introduction

Une difficulté essentielle dans les systèmes de dialogue à forte composante graphique concerne le problème du positionnement d'objets à l'écran qui requiert une bonne modélisation de l'espace et des actions [Schang et Romary, 1994b]. À notre connaissance, il n'existe pas de système qui réalise ce type d'opération tout en effectuant une compréhension fine des mécanismes sémantiques sous-jacents. Des systèmes comme Compèrobot [Andrès, 1993] ou les travaux de Arnold et Lebrun [1992] n'étudient pas le positionnement en tant que tel mais s'en servent comme outil d'étude des mécanismes plus généraux relatifs à la modélisation de l'espace en langage naturel. Trop souvent, on a pensé que des énoncés de positionnement se réduisaient au prototype : "*mets ça ici*". Il n'en est rien comme le démontre l'expérience de type Magicien d'Oz mise en œuvre par Mignot [1993].

Le problème du positionnement semble très important car il repose, pour un bonne part, sur une compréhension fine des mécanismes de référence et plus généralement des intentions du locuteur afin de déterminer exactement l'endroit du positionnement. Bien des problèmes liés à la référence sont sous-jacents aux énoncés de positionnement. Ainsi, prenons la phrase : "*rapproche le canapé de la cheminée*" recueillie dans le corpus de

Mignot. Il faut que la référence au canapé ne soit pas ambiguë même si un autre canapé se trouve à l'extérieur de la pièce. Les requêtes du type : “*rapprocher A de B*” jouent dans un espace où se trouvent déjà A et B et cela contraint la référence à A dans le cas où la référence à B n'est pas ambiguë. Dans ce cas précis, “*la cheminée*” devient centrale puisqu'un espace incluant ce site devient pertinent pour l'identification de la cible. Cet espace incluant n'est autre que le cadre de référence. On peut envisager l'hypothèse inverse pour la phrase : “*mets le canapé dans le salon*” où au contraire, seuls les meubles à l'extérieur de la pièce devraient être concernés pour en extraire le canapé. Ces deux exemples démontrent que les énoncés du type “*mets A prép. B*” ou “*rapproche A prép. B*” ne fonctionnent pas de manière purement compositionnelle suivant les trois étapes suivantes :

- recherche de A,
- recherche de B,
- placement de A par rapport à B tel que A se trouve “prép B”.

Au contraire, on assiste à une très forte imbrication de ces trois étapes et la non compositionnalité des requêtes de positionnement nous conduisent à affirmer que le problème du positionnement dépasse en complexité celui de la référence.

Cette section fera tout d'abord l'objet de la présentation des travaux de Romary [1993] qui nous permettront d'étudier en détail les requêtes qui font intervenir le déictique “*ici*”. Nous esquisserons ensuite, sans entrer dans les détails faute d'une réflexion suffisamment aboutie, ce que pourrait être un modèle de positionnement fondé sur la notion de cadre de référence.

4.8.2 Vers un positionnement plus “cognitif”, le cas de “*ici*”

Romary [1993] démontre que bien trop souvent on a associé “*ici*” à une portion élémentaire de l'espace, en l'occurrence un point M de coordonnées (x,y) dans le plan de désignation³⁶. Réduire le lieu à un point a une conséquence immédiate sur les énoncés de positionnement : un objet ne pourra être placé sur l'écran qu'à la condition qu'un point de celui-ci soit associé directement au lieu de désignation.

Cependant, dès que la tâche se complexifie un peu, un certain nombre de problèmes apparaissent. Le lieu correspondant au geste ne peut plus être simplement recherché comme un point précis de l'espace, au risque de conduire à des interprétations erronées. Ainsi, si on considère une tâche d'aménagement intérieur, l'énoncé “*mets de la moquette ici*” associé à un geste ne peut s'interpréter comme une demande limitant le positionnement de la moquette au point précis où il y a eu geste³⁷. Il est donc nécessaire d'envisager une certaine extension au lieu associé à “*ici*”.

Ainsi, dans une circonstance similaire au cas précédent, l'interprétation de l'énoncé “*mets trois prises de courant ici*” doit conduire à ce que trois prises soient globalement

36. Cette analyse est partagée par de nombreux auteurs comme en témoignent un certain nombre d'articles des actes des journées IHM 92. On peut remarquer qu'il y a alors un petit côté arbitraire à subordonner la description de l'espace à la définition (en pixel/pouces) de l'écran.

37. Si on imagine de plus qu'il s'agit de carrés de moquette, il n'est peut-être pas judicieux de les empiler en un point précis de la pièce.

associées à la pièce considérée, pour ultérieurement les voir apparaître sur les murs de celle-ci. La prise en compte d'aspects fonctionnels semble donc là encore inéluctable. L'auteur aboutit aux deux conclusions suivantes :

- “*ici*” en association à un geste, active un lieu qui semble être autre chose qu'un simple point. Ce lieu possède une certaine étendue et même une certaine structure (ce peut être une pièce dans un appartement). Dans des tâches de positionnement, on associera globalement l'objet au lieu (exemple : les prises à la pièce désignée),
- dans un deuxième temps, le positionnement final de l'objet par rapport au lieu va dépendre de contraintes inhérentes à la tâche et ne paraît pas être du ressort de la sémantique de l'énoncé. Il faudra donc posséder une description de la manière dont on place de la moquette ou des prises de courant dans une pièce (i.e. les prises de courant sur les murs).

Romary démontre ensuite la nécessité d'introduire la notion de pavage qui consiste en fait à découper la portion d'espace cible en entités élémentaires. Ce pavage ne correspond bien sûr pas à une représentation unique de l'espace, mais à un point de vue possible que le locuteur cherchera à faire partager à l'auditeur dans le discours. Ainsi, un appartement pourra être vu comme un ensemble de pièces, un jeu d'échec comme un ensemble de cases, etc.³⁸

La notion de point de vue associée à un pavage est particulièrement importante puisqu'elle permet de définir différentes configurations spatiales pour une même entité. Supposons ainsi une table, celle-ci peut-être vue dans le cadre d'un pavage représentant une dimension verticale pour exprimer des relations du type “*la lampe se trouve au-dessus de la table*” ou en opposition avec d'autres surfaces planes “*ne mets pas le livre sur la table, pose le sur l'étagère*”. La définition d'un pavage de l'espace sur lequel va s'opérer l'interprétation de “*ici*” permet de donner une certaine granularité à la précision de la référence spatiale associée.

Finalement, l'auteur énonce la règle suivante pour “*ici*” :

«le déictique “ici” réalise un filtrage sur une structure de pavage de l'espace pour isoler l'un des éléments de cette structure qui se trouve directement mis en évidence par le locuteur (i.e. intentionnellement), soit par la simple situation d'énonciation (position du locuteur ou du message), soit par un geste lorsque l'unité de pavage est relativement fine.»

Outre l'originalité de la notion de pavage, nous retiendrons en particulier de cette approche qu'elle plaide là encore en faveur d'une prise en compte accrue des phénomènes fonctionnels faute de quoi, le modèle développé s'avérerait en partie incorrect.

On pourra remarquer que la notion de pavage que propose Romary revient en fait à parler là encore d'alternatives. En effet, un énoncé du type : “*mets le fauteuil ici*” revient en fait à isoler une portion particulière de l'espace où placer le fauteuil par opposition à d'autres positions, les alternatives.

38. Entre deux unités, on peut avoir besoin d'un espace de séparation (pour éviter les cas limites).

4.8.3 Quelques pistes pour formaliser les énoncés de positionnement

Nous proposons d'étudier spécifiquement les énoncés du type "*mettre*³⁹ *A prép B*" qui sont apparus majoritairement dans le corpus de Mignot. Tout comme pour les énoncés purement référentiels, nous laisserons de côté les énoncés qui font appel à un ou plusieurs gestes de désignation.

Si les syntagmes A et B sont des groupes nominaux pouvant contenir à leur tour des informations spatiales, on pourra remarquer que ces syntagmes seront prioritairement composés de référents non spatiaux (phrases (161) et (162)) même si le cas est possible (phrase (163)). Le fait que la composition de ces syntagmes soit plus rarement spatiale provient sans doute des imbrications spatiales qui deviennent très vite lourdes à gérer d'un point de vue cognitif. En effet, ces imbrications imposent à l'utilisateur de trop fréquentes reconsidérations spatiales de la scène affichée. À ce titre, il suffit de prendre l'exemple de la phrase (164) qui n'est pas naturelle malgré la pause que marque la virgule entre l'énoncé de position initiale et l'énoncé de position finale. Il semble au contraire qu'il faille laisser un certain temps de latence à l'utilisateur afin d'assimiler le cadre de référence courant avant qu'il ne puisse reconsidérer un nouveau cadre de référence (ce point nécessiterait bien sûr une validation psycholinguistique). Pour illustrer cette remarque, il suffit de considérer les phrases (165) et (166) qui semblent bien plus "digestes" que la phrase (164) car elles laissent plus de temps à l'utilisateur pour construire un cadre de référence initial où il va rechercher la(les) cible(s) et un cadre de référence final centré sur le(s) site(s) où il va pouvoir placer la(les) cible(s).

Installer le canapé face à la cheminée. (161)

Placer le fauteuil perpendiculairement au canapé. (162)

Mets le fauteuil de droite en bas à gauche. (163)

? *Mets le fauteuil en bas à gauche, en haut à droite.* (164)

Mets le fauteuil qui se trouve en bas à gauche, en haut à droite. (165)

Prends le fauteuil qui se trouve en bas à gauche et mets le en haut à droite. (166)

Au vu de ces quelques remarques, nous envisageons le positionnement comme le déplacement d'un ou plusieurs objets d'un cadre de référence initial vers un cadre de référence final. Le modèle abstrait de positionnement pour les phrases du type "*mettre A prép. B*" qui est représenté à la figure 4.42 indique que l'on va déplacer une ou plusieurs cibles appartenant au focus initial vers un focus final. Les contributions que sont l'analyse syntaxico-sémantique, les connaissances provenant de l'application, la perception visuelle, le lexique et l'historique du dialogue seront autant d'étapes d'affinage qui vont progressivement nous conduire au calcul d'un cadre de référence initial et d'un cadre de référence

39. Notons que "*mettre*" peut être remplacé par tout verbe possédant une sémantique voisine comme "*placer*", "*installer*", etc.

final dans lequel le ou les objets seront positionnés.

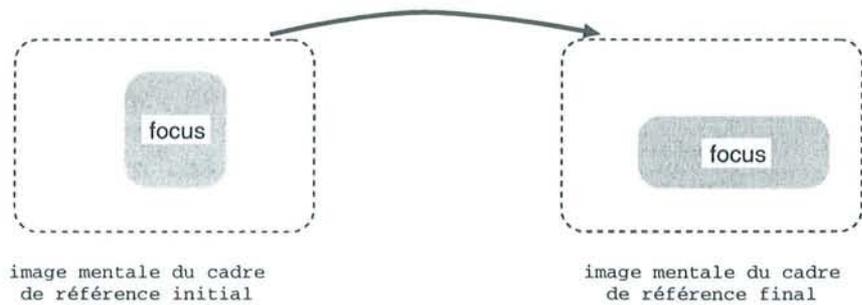


FIG. 4.42 – *Modèle abstrait pour le positionnement.*

a) Différences entre les énoncés de positionnement et les énoncés purement référentiels

Les différences que l'on peut faire entre ces deux types d'énoncés rejoignent la distinction que nous avons opérée au paragraphe 3.5 du précédent chapitre entre un énoncé descriptif et un énoncé purement référentiel. Les énoncés de positionnement imposent davantage de contraintes sur les référents que les énoncés purement référentiels. À titre d'illustration, il suffit de considérer la figure 4.43 qui démontre qu'un énoncé de positionnement est beaucoup plus précis qu'un énoncé référentiel. En effet, la requête "efface le canapé qui se trouve devant la cheminée" est tout à fait valide à la figure 4.43.(a) alors que la phrase "mets le canapé devant la cheminée" à la figure 4.43.(b) impose d'aboutir à un alignement précis entre le canapé et la cheminée.

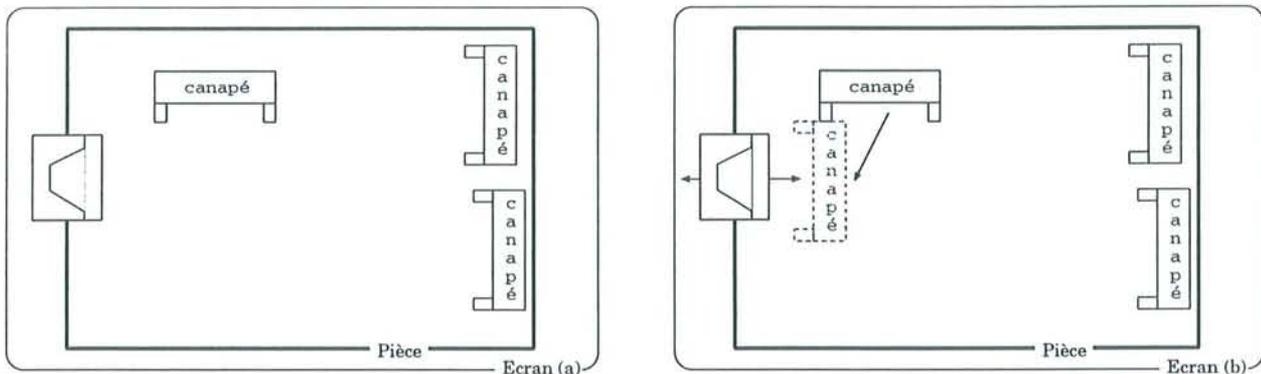


FIG. 4.43 – *Nuance entre le positionnement et la référence "pure".*

Par ailleurs, on peut parfaitement faire référence à un canapé qui tourne le dos à la cheminée, c'est-à-dire que sa position en terme d'angle de rotation est tout à fait quelconque. En revanche, pour un énoncé de positionnement, le fait qu'un objet puisse lui tourner le dos ou qu'il se trouve face à lui est tout à fait essentiel.

Ces contraintes supplémentaires par rapport aux énoncés purement référentiels nous incitent à ajouter au modèle précédemment proposé un module intitulé “connaissances sur le monde” que l’on peut intégrer aux contributions provenant de l’application. Les connaissances sur le monde serviront donc essentiellement à affiner davantage l’emplacement exact de la(des) entité(s) à positionner. Ainsi pour la requête (1) du corpus de Mignot (p. 204) où l’utilisateur souhaite placer le canapé en face de la cheminée, il faut qu’il soit placé suffisamment près de cette dernière afin de profiter de la chaleur, mais pas trop près non plus afin de laisser assez de place autour du canapé pour pouvoir bien circuler, etc. Par ailleurs, en termes de rotation, une fois la position trouvée, il convient d’orienter convenablement le canapé. On pourra remarquer que si les connaissances sur le monde semblent difficiles à mettre en œuvre, de notre point de vue, elles sont à même de rendre le système de positionnement vraiment efficace.

Au terme de l’élaboration du cadre de référence initial et du cadre de référence final, on doit déboucher sur un focus initial et un focus final suffisamment précis pour qu’il n’y ait pas d’ambiguïtés dans le choix des éléments à positionner et de la(des) position(s) qui sera(seront) occupée(s) par cette(ces) entité(s). Dans le cas contraire, l’ordinateur sera amené à demander plus d’information. C’est le cas pour la requête (167) du corpus de Mignot où il y a ambiguïté au niveau de l’ensemble initial car deux choix de canapés sont possibles. On pourrait croire de prime abord que l’utilisateur ne respecte pas les maximes de Grice (celle d’être suffisamment précis en particulier), en fait il n’en est rien, le locuteur, dans sa perspective n’a même pas prêté attention au canapé clair qui se trouve en dessous du canapé grisé, pour lui, seul ce dernier a été pris en compte.

Installez le canapé face à la cheminée.

(167)

b) Autres difficultés liées au problème du positionnement

Outre les difficultés inhérentes à la gestion de deux cadres de référence : le cadre de référence initial et le cadre de référence final, les éventuelles contraintes de disjonction spatiale qui pèsent sur ces derniers complexifient la tâche de modélisation. À titre d’illustration, il suffit de considérer l’énoncé “*installez le canapé face à la cheminée*”. Dans ce cas, le cadre de référence initial et le cadre de référence final sont disjoints, le premier faisant partie de l’extérieur du salon tandis que le second se localise dans le salon autour du salon. Par contre, si cet énoncé était par exemple suivi de la requête “*rapprochez le de la cheminée*”, le nouveau cadre de référence initial et le nouveau cadre de référence final seraient sensiblement confondus, le verbe “*rapprocher*” opérant pour ainsi dire à cadre de référence constant.

4.8.4 Conclusion

La réflexion proposée dans ce paragraphe nous a permis de démontrer que la complexité des énoncés de positionnement dépassait celle des énoncés purement référentiels. D’autre part, la présentation des travaux de Romary concernant le traitement des énoncés faisant intervenir le déictique “*ici*” nous ont permis d’insister une fois de plus sur l’importance

de la notion d'alternatives. Enfin, nous avons pu poser les premiers jalons de ce que serait un modèle des énoncés de positionnement fondé sur la notion de cadre de référence.

4.9 Autres validations des cadres de référence

Nous avons démontré dans ce chapitre que les cadres de référence étaient sous-tendus par la notion de principe d'économie cognitive. Ce principe possède sans aucun doute quelques attraits pour un informaticien comme nous l'avons esquissé au premier chapitre. En particulier, de gros problèmes naissent de la très grande quantité de connaissances à mettre en œuvre dans un système informatique. Retrouver l'information pertinente et ne pas se noyer dans des déductions inutiles sont, bien qu'essentiels, des questions encore ouvertes. Fodor [1986] ou Haton et Haton [1990] soulignent que c'est là l'un des problèmes fondamentaux de l'intelligence artificielle à propos du *frame problem*. Les cadres de référence semblent donc pouvoir œuvrer dans le sens d'une limitation des ressources informatiques mises en jeu à un instant donné d'un discours ou d'un dialogue. Ils s'inscrivent directement dans la maxime de Hernández [1994] suivante :

« On s'accorde généralement à dire qu'il ne faut représenter que l'information nécessaire qui autorisera à un instant donné la discrimination voulue d'un objet, d'une propriété ou d'une relation dans un contexte donné. »

Si la notion de cadre de référence possède une utilité informatique, on peut se demander si elle reflète un tant soi peu le fonctionnement cognitif humain. Cette question justifie la discussion que nous allons entreprendre sur la dualité entre les modèles informatiques *versus* les modèles de la cognition humaine.

4.9.1 Modèles informatiques, modèles de la cognition humaine : un cas de divorce ?

On peut se demander, si ces représentations que l'on espère un jour voir fonctionner de façon automatique sur une machine, correspondent d'une manière ou d'une autre à des représentations mentales telles qu'un locuteur pourrait les construire dans son activité langagière propre.

Remarquons qu'il est très difficile pour ne pas dire impossible de véritablement savoir quels sont les modes de représentation que l'être humain adopte. La seule chose que l'on puisse observer (en-dehors de certaines expériences mettant en jeu des potentiels [Sergent, 1993] évoqués ou des images obtenues par RMN ou TEP⁴⁰), sont des comportements langagiers ou d'action.

Dans le cas d'un dialogue, il n'est pas véritablement important de savoir si telle ou telle représentation est analogue à une construction mentale du locuteur. Ce qui compte essentiellement c'est que la vision (et la compréhension) du monde qu'elle entraîne soit compatible avec celle du locuteur, pour aboutir à une bonne compréhension mutuelle des deux sujets en interaction et au-delà à un sentiment de naturel de la part du locuteur.

40. RMN : résonance magnétique nucléaire, TEP : tomographe par émission de positons.

Le fait de chercher à élaborer des modèles dont le fonctionnement se calquerait au mieux sur le fonctionnement cognitif humain est par ailleurs une problématique essentielle en intelligence artificielle. Ainsi, si nous abordons ce problème sous l'angle spécifique du fonctionnement du cerveau humain, on peut mener une réflexion sur la pertinence des modèles connexionnistes et se demander si le neurone doit constituer oui ou non le pilier de tout système d'intelligence artificielle. Si l'on regarde ce qui se passe dans la nature, on peut constater que le neurone apparaît avec une fidélité étonnante. Que l'on s'intéresse à la limace ou à l'homme, en passant par la mouche, le neurone est présent.

A contrario, on pourrait remarquer que le neurone ne relève après tout que d'un choix parmi d'autres, la nature n'ayant pour ainsi dire rien trouvé de mieux et ce au terme d'une longue évolution. Sur cette base, on pourrait alors chercher à fabriquer des systèmes intelligents qui s'écarteraient complètement du fonctionnement cognitif humain. On peut remarquer que l'homme y est parvenu pour nombre d'objets et matériaux qui n'existaient pas dans la nature et qui donnent des résultats très satisfaisants.

Au vu des remarques précédentes, pour le concepteur d'un système de dialogue, la question de savoir si les cadres de référence s'inspirent ou non de la cognition humaine n'aurait après tout pas grand intérêt. Il semble cependant que le fonctionnement cognitif humain, du fait de sa longue évolution, possède *a priori* une certaine efficacité, ce qui justifie que nous nous tournions à présent vers la psychologie pour essayer de donner une validité psychologique à la notion de cadre de référence.

4.9.2 Validation psychologique des cadres de référence

Piérart [1979] a réalisé une étude psycholinguistique de certains marqueurs spatiaux (“*au-dessus de*”, “*sur*”, “*en dessous de*”, “*sous*”, “*devant*”, “*derrière*”, “*autour de*”, etc). Elle a soumis ces prépositions à des adultes ainsi qu'à de jeunes locuteurs âgés de trois à dix ans. Chacun a estimé la parenté entre ces marqueurs sur une échelle possédant cinq échelons dont les extrémités sont respectivement “les prépositions donnent des indications identiques” et “les prépositions donnent des indications tout à fait différentes”. L'ensemble des résultats obtenus par des techniques d'analyse multidimensionnelle fournit un ensemble de critères de jugement des sujets que l'auteur considère comme les dimensions d'un espace psycholinguistique. Les locuteurs distinguent les prépositions suivant les axes de verticalité, de sagittalité, d'inclusivité et de latéralité.

Outre les différentes dimensions de l'espace psycholinguistique auxquelles aboutit Piérart, ses travaux illustrent également le fait que les enfants apprennent plus tardivement l'usage de certaines prépositions. En particulier, ce sont d'abord les marqueurs qui font intervenir l'intériorité/l'entourage (sous-jacents aux prépositions “*dans*”/“*autour de*”) qui sont acquis en premier. Par la suite, sont acquises les prépositions “*loin de*”/“*près de*”, “*à côté de*”, “*contre*” qui font intervenir des aspects de voisinage et de distance. Ce n'est que beaucoup plus tard que les prépositions “*devant*”/“*derrière*” sont parfaitement utilisées⁴¹. Elles nécessitent que l'enfant soit capable de maîtriser un système d'axes et de coordonnées mobiles dans l'espace ainsi que l'intégration d'un point de vue.

41. Nous insistons sur le terme de “parfaitement” car dès l'âge de 3 ans, les enfants sont capables d'utiliser la préposition “*derrière*” en lui prêtant le trait “caché” [Piérart, 1979].

À l'issue de ces quelques exemples, il semblerait que les enfants apprennent plus difficilement (plus tardivement) l'usage de certaines prépositions spatiales car la structure qu'ils doivent mettre en œuvre (qui est proche d'une structure de cadre de référence) est plus élaborée donc plus lourde à gérer d'un point de vue cognitif. Ainsi, si nous prenons l'exemple de la préposition "*derrière*", le cadre de référence qu'elle met en jeu est bien plus élaboré (du fait de l'axe de sagittalité et des problèmes de masquage) que celui que mettrait en œuvre la préposition "*près de*" par exemple.

Par ailleurs, Rock et Palmer [1991] soulignent que les partisans de la Gestalt découvrirent que certaines structures déterminent un "cadre de référence" à l'aide duquel les autres objets sont perçus. Assis dans un train, nous avons tous cru qu'il se déplaçait par rapport au train voisin, alors que c'était l'autre qui se déplaçait par rapport à nous. Une autre illustration de cette illusion apparaît lorsqu'un observateur entre dans une pièce dont les murs sont inclinés : les murs de la pièce définissent les axes verticaux et horizontaux du cadre de référence. Un chandelier posé au sol aura alors un air curieusement oblique, de même que l'observateur aura l'impression que son corps est incliné, alors que tous les deux sont parfaitement parallèles à la verticale imposée par la gravité.

Ces différents aspects plaident donc en faveur d'une certaine validité psychologique de la notion de cadre de référence. Cette validité psychologique se double également d'une validité linguistique pour le cas particulier des ambiguïtés qui peuvent surgir entre les deux locuteurs.

4.9.3 Cadres de référence et ambiguïté

Comme nous avons pu le voir au paragraphe 4.5.3, la notion de cadre de référence semble être en mesure d'expliquer certains phénomènes d'ambiguïtés référentielles spatiales qui peuvent apparaître dans le dialogue homme-machine. Nous posons donc le postulat suivant : *« les ambiguïtés référentielles spatiales surgissent entre l'homme et la machine lorsqu'ils ne partagent pas le même cadre de référence »*.

Le corpus de Mignot [1995] contient nombre de requêtes qui semblent ambiguës à l'un des deux partenaires du dialogue par le fait qu'ils ne partagent pas le même cadre de référence à un instant donné du dialogue. Il suffit de considérer la première requête de l'utilisateur qui apparaît dans le corpus : "*installez le canapé face à la cheminée*" (p. 204). Une ambiguïté surgit car l'utilisateur n'a pas du tout prêté attention au deuxième canapé. Pour lui, il n'y avait qu'un seul canapé. Ce n'est qu'au moment où le magicien d'Oz énonce "*lequel ?*" qu'il prend conscience du second canapé. Il reconsidère alors la scène suivant une opposition référentielle entre le petit et le grand canapé.

De notre côté, nous avons pu insister sur le fait que dès qu'une ambiguïté surgit au niveau d'une requête spatiale formulée par l'utilisateur, il faut que l'ordinateur change de cadre de référence jusqu'à ce que l'ambiguïté soit levée, l'ordinateur adoptant en quelque sorte le point de vue de l'utilisateur⁴². En effet, en vertu des maximes de Grice ou du prin-

42. Cette tâche de désambiguïsation n'est bien sûr pas toujours aisée comme en témoigne l'exemple :

cipe de pertinence de Sperber et Wilson, si l'utilisateur énonce une requête, cette dernière est nécessairement non ambiguë pour lui. À titre d'illustration, il suffit de reconsidérer la requête "détruis le triangle qui se trouve à droite du carré" formulée à la figure 4.44.

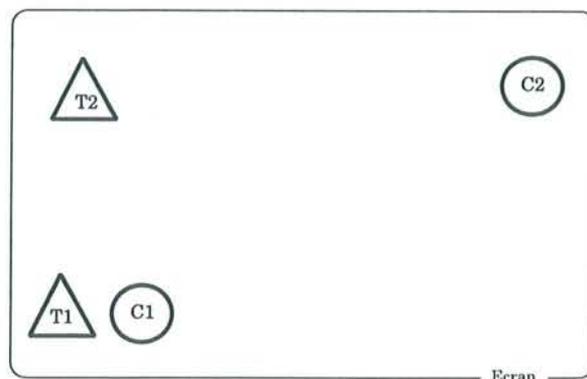


FIG. 4.44 – Cadres de référence et ambiguïté.

Si cette requête est ambiguë dans l'absolu, il faut que l'ordinateur change de cadre de référence et retienne un cadre de référence qui englobe le triangle T1 et le cercle C1. Nous rappelons que c'est effectivement ce que réalise l'algorithme que nous proposons au paragraphe 4.6. La méthodologie proposée construit un cadre de référence autour du triangle T1 ainsi qu'autour du triangle T2. Ces constructions nous conduisent à rejeter le cercle C2 car il ne fait pas partie du cadre de référence associé au triangle T2.

Bien entendu, l'utilisateur ne respectera pas toujours les maximes de Grice ou le principe de pertinence (même involontairement). Il suffit de se remémorer l'exemple "Agrandis les carrés de droite. Détruis celui de gauche" énoncé à la figure 4.45.

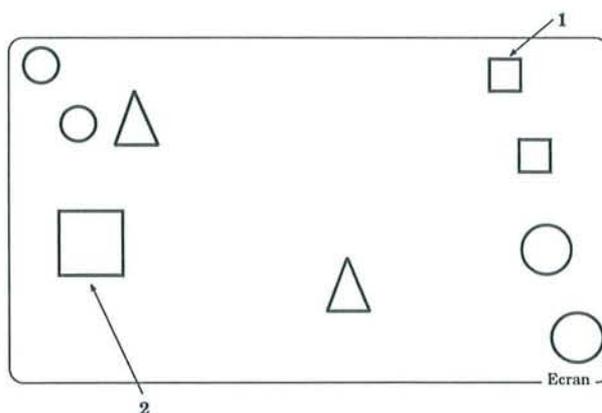


FIG. 4.45 – Un cas d'ambiguïté?

Deux cas de figure sont alors possibles. Soit l'utilisateur ne prête plus du tout attention au carré qui se trouve complètement à gauche après avoir énoncé "Agrandis les carrés de droite", dès lors il n'y a pas d'ambiguïté dans sa requête "Détruis celui de gauche" car il

"installez le canapé face à la cheminée"

souhaite détruire le carré numéro 1. Au contraire, l'utilisateur peut avoir à l'esprit une opposition référentielle entre les carrés de droite et celui de gauche. Dans ce cas, sa seconde requête vise à détruire le carré numéro 2.

De notre côté, étant donné que le modèle que nous avons proposé est capable de détecter ce type d'ambiguïté, il est nécessaire que le système invite l'utilisateur à effectuer un choix parmi l'un des deux carrés.

4.9.4 Les cadres de référence et la transitivité

Les cadres de référence semblent également en mesure d'expliquer certains phénomènes de transitivité. Examinons les phrases suivantes tirées des travaux de Vieu [1991]

Paul est dans la maison (168)

La maison est dans l'île (169)

$(168) \wedge (169) \implies \text{Paul est dans l'île.}$ (170)

Paul est dans l'île (171)

L'île est dans la mer (172)

$(171) \wedge (172) \implies \text{Paul est dans la mer}$ (173) (Faux)

De notre point de vue, il n'est pas si sûr de pouvoir déduire que "*Paul est dans l'île*" à partir des phrases (168) et (169). Ainsi "*Paul est dans la maison*" prend sa signification en fonction des alternatives, si Paul est soit dans la maison (il se repose) soit dans l'île (et non pas sur le continent), Paul n'est surtout pas dans l'île (sauf si on change de cadre de référence à nouveau pour se demander si Paul est dans l'île ou le continent).

De même pour (173), lorsqu'on dit que "*Paul est dans l'île*" par opposition à "*Paul est sur le continent*", on peut très bien imaginer qu'un typhon se serait abattu sur cette île emportant Paul au large, rendant alors envisageable le fait que Paul se trouve dans la mer!

La nature des cadres de référence et plus particulièrement leur fonctionnement à base d'alternatives semble donc à même de pouvoir décrire certains aspects de la transitivité spatiale en langage naturel.

4.9.5 Liens avec le temps

Le temps joue un rôle très important dans tout système de dialogue homme-machine ne serait-ce qu'au niveau de l'évolutivité des référents [Gaiffe et Romary, 1995]. Pierrel et Romary [1997] soulignent que la plupart des analyses qui s'intéressent au temps, dans une perspective de modélisation informatique, ne conçoivent que trop rarement ce dernier comme résultant d'un processus référentiel. En effet, le temps, perçu comme linéaire par les systèmes informatiques s'éloigne de la conception subjective que nous en avons, nous autres humains.

Par ailleurs, la plupart des analyses actuelles n'accordent qu'une faible place au contexte d'interprétation. Ce contexte d'interprétation fait directement le lien avec notre notion de cadre de référence. En effet, il semble naturel de penser qu'un énoncé comme

“*tous les étudiants danseront*” ne peut être interprété s’il n’a pas été posé au préalable un cadre temporel où le futur peut effectivement prendre sa valeur par rapport au présent.

On pourra trouver dans Romary [1991] les bases d’une méthode de représentation intégrée d’informations spatiales et temporelles qui aboutit à la définition d’un *univers d’objets* qui permet de suivre les différents états et événements liés à un objet décrit dans un discours en langage naturel. La méthodologie proposée s’appuie sur les travaux antérieurs de Romary [1989] relatifs aux zones temporelles (nous ferons une brève présentation de la notion de zone temporelle au chapitre 5).

4.10 Conclusion

Ce chapitre s’est articulé suivant cinq volets. Le premier nous a permis de montrer la nécessité d’introduire la notion de cadre pour laquelle nous avons proposé une définition et décrit ses caractéristiques. Le second volet s’est intéressé à l’instanciation de la notion générale de cadre pour le problème spécifique de la référence spatiale ce qui nous a conduit à définir, caractériser et situer par rapport aux autres approches la notion de cadre de référence. Le troisième volet s’est situé dans une perspective d’implémentation. Nous avons pu étudier en détail quelles étaient les connaissances à mettre en œuvre pour réaliser correctement l’interprétation d’une référence spatiale au sein d’un système de dialogue homme-machine à forte composante graphique. Le quatrième volet a posé le problème de l’adaptabilité de la notion de cadre de référence au problème spécifique du positionnement ce qui nous a amené à poser les premiers jalons d’un tel modèle. Le dernier volet de ce chapitre nous a permis de donner une validité psychologique à la notion de cadre de référence.

Parmi les perspectives d’extension du modèle proposé, il semble nécessaire de lui adjoindre à court terme une composante gestuelle dont les principes de fonctionnement sous-jacents semblent présenter une très forte connexité avec certains aspects mis en œuvre au sein des cadres de référence : en effet, une désignation gestuelle sert essentiellement à attirer l’attention de l’interlocuteur sur un sous-espace du champ perceptif.

Par ailleurs, l’amorce de la réflexion qui a été présentée concernant les aspects transitifs plaide en faveur de l’utilisation de la notion de cadre de référence afin d’expliquer et modéliser certains phénomènes de transitivité.

À plus long terme, nous avons pu esquisser dans ce chapitre qu’une certaine vision de temps n’était pas incompatible avec les principes que nous avons énoncés concernant les cadres de référence, ce qui plaide en faveur d’une extension possible de ces derniers à la prise en compte de certains aspects temporels.

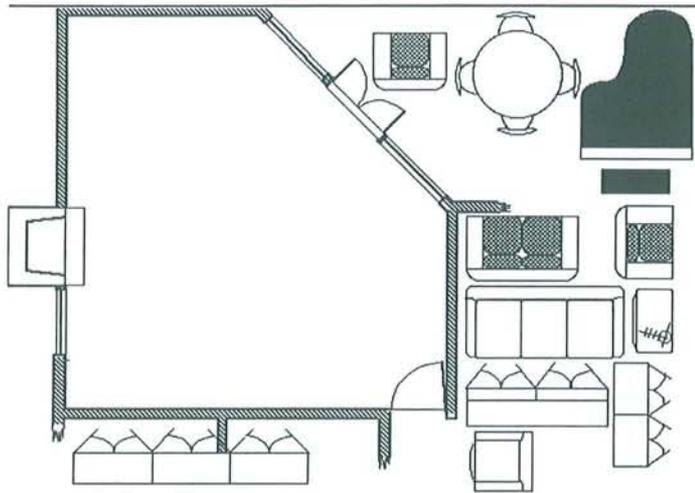


FIG. 4.46 – Visualisation d'un salon pour une application d'aménagement.

Voici l'extrait du corpus de dialogue tiré de [Mignot *et al.*, 1993] :

<i>Locuteur</i>	<i>Magicien</i>
(1) installez le canapé face à la cheminée	(a) lequel?
(2) le plus petit	(b) compris
(3) installez un fauteuil	(c) (action) ici?
(4) celui-ci (geste)	(d) lequel?
(5) perpendiculairement au canapé	(e) compris
(6) oui, ici exactement (geste)	(f) où exactement?
[...]	(g) (action) ici?
(7) placez à présent l'autre fauteuil perpendiculairement au canapé	(h) compris
(8) ici (geste)	(i) où exactement?
[...]	(j) compris
(9) installez le piano dans l'angle près de la porte	(k) (action) êtes-vous satisfait?
(10) rapprochez le de la cheminée	(l) compris (action)
[...]	

Chapitre 5

Intégration d'une composante spatiale dans un système de dialogue

Sommaire

5.1	Introduction	205
5.2	Les systèmes de dialogue	206
5.2.1	Fonctions à mettre en œuvre dans un système de dialogue . . .	206
5.2.2	Les connaissances à mettre en œuvre dans un système de dialogue	208
5.2.3	Bilan	209
5.3	Intégration d'une composante spatiale dans un système de dialogue	209
5.3.1	“Un fondement solide” : le modèle de CAO	210
5.3.2	Bilan de ce modèle	213
5.4	La maquette MultiDial	215
5.5	Conclusion	222

5.1 Introduction

L'intérêt de ce chapitre réside dans les perspectives d'exploitation informatique de l'étude qui a été réalisée dans cette thèse, notre but étant à terme d'offrir à l'utilisateur la possibilité d'exprimer ses références spatiales en utilisant la parole. Notre démarche va consister à définir les connaissances et les traitements à intégrer dans un système de dialogue afin de permettre la gestion d'énoncés spatiaux. Notre but n'est pas de décrire précisément le fonctionnement d'un système de dialogue mais, simplement, d'en décrire les différents aspects en vue d'une intégration future de la notion de cadre de référence.

Ce chapitre s'articulera en trois volets. Nous préciserons tout d'abord ce que l'on entend classiquement par système de dialogue en présentant tout d'abord leurs différentes fonctionnalités puis les connaissances sur lesquels ils reposent. Dans un second temps,

nous présenterons l'état de la maquette réalisée, qui s'inspire de nombreux concepts issus de la CAO¹. Le dernier volet de ce chapitre fera l'objet de la présentation de la maquette MultiDial au sein de laquelle nous comptons implémenter notre système.

5.2 Les systèmes de dialogue

Cette partie va nous permettre de revenir sur les fonctions et connaissances qui sous-tendent la réalisation d'un système de dialogue. Pour mieux situer l'ensemble de ces fonctions et connaissances, il ne faut pas perdre de vue l'objectif premier d'une interface de dialogue homme-machine qui consiste à minimiser l'effort d'adaptation de l'utilisateur à la machine. Cette minimisation permet alors à l'utilisateur de se concentrer bien plus sur le fond de son problème plutôt que sur sa forme qui doit être décrite à l'ordinateur. De manière générale, on pourrait définir un système de dialogue comme étant un partenaire compétent et coopératif dont la raison d'être est d'aider un utilisateur à mener à bien une tâche particulière dans le cadre d'une application donnée.

5.2.1 Fonctions à mettre en œuvre dans un système de dialogue

Les principales fonctions à mettre en œuvre dans un système de dialogue, en plus des composantes morphologique, phonétique, phonologique, lexicale, syntaxique et sémantique sont les suivantes d'après [Carbonell *et al.*, 1985] et [Carré *et al.*, 1991] :

a) La gestion du canal de communication

Outre la prise en compte du début et de la fin du dialogue, cette fonction produit des énoncés visant à entretenir le dialogue. En particulier, il faut que la machine produise des messages d'attente lorsque sa réponse à la requête de l'utilisateur est relativement longue. De même, il faut relancer l'utilisateur lorsque ce dernier tarde à répondre ou tarde à exprimer une nouvelle requête.

b) L'interprétation contextuelle

Il s'agit ici d'intégrer la représentation syntaxico-sémantique provenant du dernier énoncé du locuteur dans le contexte du dialogue. Cette fonction doit distinguer les énoncés qui portent sur la tâche de ceux portant sur le dialogue tout en les interprétant.

c) La gestion du dialogue

La gestion du dialogue s'occupe de gérer les cas de contestation, les demandes de reformulation, les prises en compte des accusés de réception ou de satisfaction.

1. CAO : Conception Assistée par Ordinateur.

d) La génération

Fonction concomitante à la reconnaissance, la génération produit les questions ou les réponses du système en ayant recours à un modèle de synthèse de la parole. On remarquera qu'une synthèse ou une restitution de la parole n'est pas toujours idéale. En particulier, dans certaines applications où l'utilisateur est déjà à l'écoute d'un signal, un retour vocal n'est pas envisageable².

e) Les mécanismes de raisonnement

Cette fonction gère les énoncés de l'utilisateur portant sur la tâche. En se fondant sur l'historique du dialogue, le modèle de la tâche, l'univers de la tâche et l'ensemble des solutions partielles provenant de la mémoire à court terme³, les mécanismes de raisonnement permettent d'effectuer les opérations suivantes :

- la détection des contradictions internes au système et la détection des incohérences externes qui proviennent de l'inadéquation d'une demande de l'usager en regard des compétences du système,
- la résolution des références anaphoriques. C'est à ce niveau que se situe l'une des contributions de cette thèse qui cherche à améliorer la composante de calcul des référents spatiaux,
- le traitement des énoncés elliptiques,
- la détection et le suivi du thème du dialogue ou du sous-dialogue en cours.

Deux cas peuvent se présenter au terme de l'analyse du module de raisonnement. Soit la demande de l'utilisateur est suffisamment explicite auquel cas, le système peut provoquer l'action correspondant à la demande de l'usager. Soit la demande de l'utilisateur est incomplète, le système cherche alors à compléter cette demande en activant le sous-but à valider ou à atteindre ce qui conditionnera la suite du dialogue.

f) La prédiction

Cette fonction cherche à prédire la future intervention de l'utilisateur en se fondant sur l'échange qui a eu lieu entre le système et l'utilisateur. La prédiction peut aboutir à préciser l'énoncé qui va suivre en fournissant :

- son type possible,
- les structures syntaxico-sémantiques les plus probables qui seront utilisées,
- un sous-lexique et un ensemble de mots clés qui seront autant de points d'ancrage pour la suite de l'analyse. Les mots clés auront en particulier un rôle décisif à jouer dans la reconnaissance de la parole.

Pour mettre en œuvre les différentes fonctions que nous venons d'énumérer, le système de dialogue doit s'appuyer sur un ensemble de connaissances que nous allons détailler à présent.

2. Voir à ce titre les travaux de Souvay [1992] portant sur la réalisation d'une plate-forme informatique capable de gérer une console sonar pour laquelle une restitution d'un message vocal en plus du signal provenant du sonar n'est pas souhaitable.

3. Les notions d'historique du dialogue, de modèle de la tâche, d'univers de la tâche et de mémoire à court terme seront décrits à la section suivante.

5.2.2 Les connaissances à mettre en œuvre dans un système de dialogue

Selon Pierrel [1987], deux types de connaissances sont nécessaires afin de mettre en œuvre un véritable système de dialogue. Il s'agit en particulier des connaissances statiques et dynamiques.

a) Les connaissances statiques

a.1) Le modèle du langage

Ce modèle fournit, sur la base de connaissances lexicales, syntaxiques et sémantiques une description des énoncés possibles. Il sert tant du point de vue de la génération que de la reconnaissance des phrases du locuteur.

a.2) Le modèle de la tâche

Ce modèle fournit au système une représentation du monde de référence, des différents objets manipulés et des relations qu'ils entretiennent entre eux. On distingue dans ce modèle deux aspects :

- une définition des objets et des relations entre les objets sur la base d'une représentation sémantico-pragmatique,
- une définition en termes de buts et de sous-buts qui spécifient les chemins d'accès aux données, fonctions ou renseignements de l'application.

On notera que c'est dans ce modèle que sont définies les expressions de généralité, ou valeurs par défaut qui sont nécessaires au traitement des énoncés. En particulier, à l'oral plus encore qu'à l'écrit, les énoncés sont très concis et ne précisent bien souvent qu'un ensemble minimal de paramètres.

a.3) Le modèle du dialogue

Ce modèle décrit les diverses situations de dialogue possibles liées à l'application considérée. Il peut s'agir des demandes de répétition, des demandes de confirmation, de contestation, etc. Ce modèle fournit les stratégies de gestion appliquées à chaque cas.

a.4) Le modèle de l'utilisateur

Ce modèle se fonde sur les particularités du locuteur qui relèvent des caractéristiques acoustico-phonétiques (définition de règles phonologiques et prosodiques) et des caractéristiques du vocabulaire qui sont utilisées par le locuteur. Par ailleurs, ce modèle prend en charge la gestion des droits d'accès de l'utilisateur concernant les opérations sensibles.

Voyons à présent les connaissances dynamiques qui évoluent au fur et à mesure que se déroule le dialogue.

b) Les connaissances dynamiques

Ces connaissances qui varient fréquemment avec le temps sont liées d'une part au côté évolutif de la tâche et d'autre part au dialogue en cours. Elles se déclinent suivant trois aspects :

b.1) L'univers de la tâche

En complément du modèle de la tâche, ce modèle fournit une description de l'état courant de la tâche : quels sont les objets de la tâche en cours, quel est leur état, à quelle position de l'écran sont-ils affichés, quelle est leur forme, leur taille, leur couleur, etc. Ce modèle est décisif pour interpréter les énoncés lorsque la tâche évolue. Nous proposerons au paragraphe 5.3 une représentation possible des objets de la tâche pour un système d'affichage graphique.

b.2) L'historique du dialogue ou mémoire à long terme

L'historique du dialogue permet de garder en mémoire la forme langagière des différentes interventions des interlocuteurs ainsi que l'organisation logique du dialogue. Ce modèle a donc un rôle essentiel à jouer pour traiter les incohérences (revoir les incohérences internes et externes du paragraphe précédent) mais aussi les tournures elliptiques et anaphoriques. Nous verrons au paragraphe 5.4 comment un tel historique a été mis en œuvre au sein de la maquette MultiDial sur la base des zones temporelles.

b.3) La mémoire à court terme

En se fondant sur les niveaux d'analyse acoustique, phonétique et lexicale du dernier énoncé, cette mémoire permet de gérer les diverses interprétations partielles possibles lors d'une phase élémentaire de dialogue.

5.2.3 Bilan

Comme nous l'avons souligné dans cette section, notre contribution porte au niveau des mécanismes de raisonnement qui apparaissent dans un système de dialogue, notamment au niveau de la résolution des anaphores spatiales. Par ailleurs, nous avons également mené en parallèle une réflexion sur la représentation des objets de la tâche que nous allons décrire à présent.

5.3 Intégration d'une composante spatiale dans un système de dialogue

Nous avons poursuivi deux objectifs principaux lors de la réalisation de la maquette de CAO que nous allons présenter ci-après. L'objectif premier a consisté en l'élaboration d'un modèle de représentation et d'affichage d'objets graphiques qui permette de tester les

algorithmes de calcul des référents spatiaux et ce indépendamment de tout autre système. En particulier, le système devait autoriser la simulation de représentations 2D, 2D1/2 ou 3D. Le second objectif procédait de l'utilisation future d'un tel modèle au sein d'un module de dialogue qui possède, ainsi que nous le décrirons pour le système MultiDial, sa propre représentation des objets de la tâche.

5.3.1 “Un fondement solide” : le modèle de CAO

Notre principale préoccupation lors de la réalisation de la plate-forme spatiale a été de la rendre la plus évolutive possible. En particulier, les soucis de portabilité et les possibilités d'interfaçage semblent déterminantes. L'évolutivité nous a poussé à rechercher au sein du domaine de la CAO un modèle qui offrait cette potentialité. Notre choix a porté sur un modèle de CAO intitulé “modèle de CAO-3D à facettes planes orientées”. Si ce type de modèle autorise la visualisation de solides en trois dimensions sous n'importe quel angle de vue, il permet également d'afficher des objets graphiques en deux dimensions. La construction des objets s'effectue de manière incrémentale suivant les opérations suivantes :

- CRÉER-POINT,
- CRÉER-ARÊTE,
- CRÉER-FACE,
- CRÉER-SOLIDE.

On dispose bien entendu des opérations concomitantes de destruction des objets de manière descendante en ayant recours aux fonctions : DÉTRUIRE-SOLIDE, DÉTRUIRE-FACE, DÉTRUIRE-ARÊTE ET DÉTRUIRE-POINT. Pour une plus grande souplesse d'utilisation de notre modèle, nous avons créé les opérateurs de haut niveau suivants : CRÉER-CARRÉ, CRÉER-RECTANGLE, CRÉER-TRIANGLE, CRÉER-CERCLE, CRÉER-CUBE, CRÉER-PYRAMIDE, CRÉER-PARALLÉLÉPIPÈDE, CRÉER-PYRAMIDE, CRÉER-SPHÈRE. Ces opérateurs sont par ailleurs très utiles car les opérations de création et de destruction des différentes composantes d'un objet (faces, arêtes, points) deviennent très rapidement fastidieuses à manipuler. À titre d'exemple, il suffit de prendre le cas de la construction d'une sphère qui requiert la création de nombreuses facettes afin que le phénomène de facettisation ne soit pas trop sensible à l'affichage.

Les facettes des objets ont par ailleurs un rôle essentiel à jouer dans ce type de modélisation. En particulier elles sont orientées selon la convention “matière à gauche” c'est-à-dire qu'en parcourant les arêtes de la face perpendiculairement à cette dernière, la matière se trouve à gauche (figure 5.1). Le fait de pouvoir situer précisément où se situe la matière est très important pour les modes de rendu réaliste où il faut peindre les faces. En effet, dans certains cas, il est nécessaire de réaliser des opérations d'intersection de faces lorsqu'un solide en cache partiellement un autre, puis reconstruire les facettes. Cette dernière opération ne peut se faire sans savoir précisément où se trouve la matière [Dony, 1985].

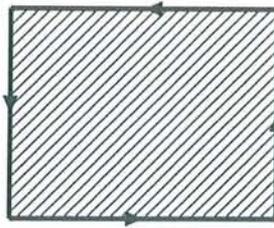


FIG. 5.1 – La convention “matière à gauche”.

En nous fondant autant que faire se peut sur les principes sous-jacents aux LGB⁴, l'accès aux objets construits n'est rendu possible qu'à l'aide d'opérateurs de haut niveau comme :

- AFFICHE-ENTITÉ (liste-entités, mode-affichage⁵),
- EFFACE-ENTITÉ (liste-entités),
- TRANSLATE-ENTITÉ (dx, dy, dz),
- ROTATION-ENTITÉ (teta-x, teta-y, teta-z).

Par ailleurs, toutes les opérations “classiques” de visualisation sont possibles à partir de la scène d'objets considérée : le zoom avant, arrière, le changement de la position du point de vue ainsi que la direction du regard. Pour le rendu réaliste de la scène affichée, comme le préconise Gardan [1987], l'affichage d'une scène se décompose en trois étapes (figure 5.2) :

- le passage des coordonnées 3D réelles du repère (i, j, k) aux coordonnées 3D transférées dans le repère de l'observateur (i', j', k'),
- le passage de ces dernières aux coordonnées 2D projetées (appelées coordonnées de la fenêtre),
- le passage des coordonnées de la fenêtre (à valeur réelles) aux coordonnées clôture (à valeurs entières). Ce dernier passage autorise un affichage de la scène dans n'importe quelle portion de l'écran. Pour un exposé détaillé des formules de transformation d'une coordonnée à l'autre voir [Gardan, 1987] ou [Dony, 1985].

Il est très important de travailler dès que possible dans l'espace clôture car cet espace nous permet d'optimiser sensiblement les opérations d'effacement, de translation, de rotation et ce, pour au moins deux raisons. La première raison procède du fait que seules les coordonnées de la clôture sont du type entier ce qui accélère sensiblement les calculs qui y sont faits. D'autre part, si l'on prend l'exemple d'un solide qui subirait une translation à l'écran, il paraît judicieux de ne réafficher, à un instant donné du déplacement que le ou les solide(s) dont l'affichage a pu être affecté par cette translation. Pour ce faire, nous avons préféré tester les chevauchements éventuels au niveau des enveloppes 2D plutôt que

4. LGB : Langages Graphiques de Base, il s'agit de modèles réalisant des interfaces de conception graphiques “propres” en se fondant sur une construction modulaire des entités de la tâche, voir [Gardan, 1987] pour un exposé détaillé.

5. Les quatre modes autorisés (fil de fer, fil de fer sans faces arrières, facettes planes + éclairage constant + élimination des facettes arrières + tracé des arêtes, rendu réaliste) seront explicités par la suite.

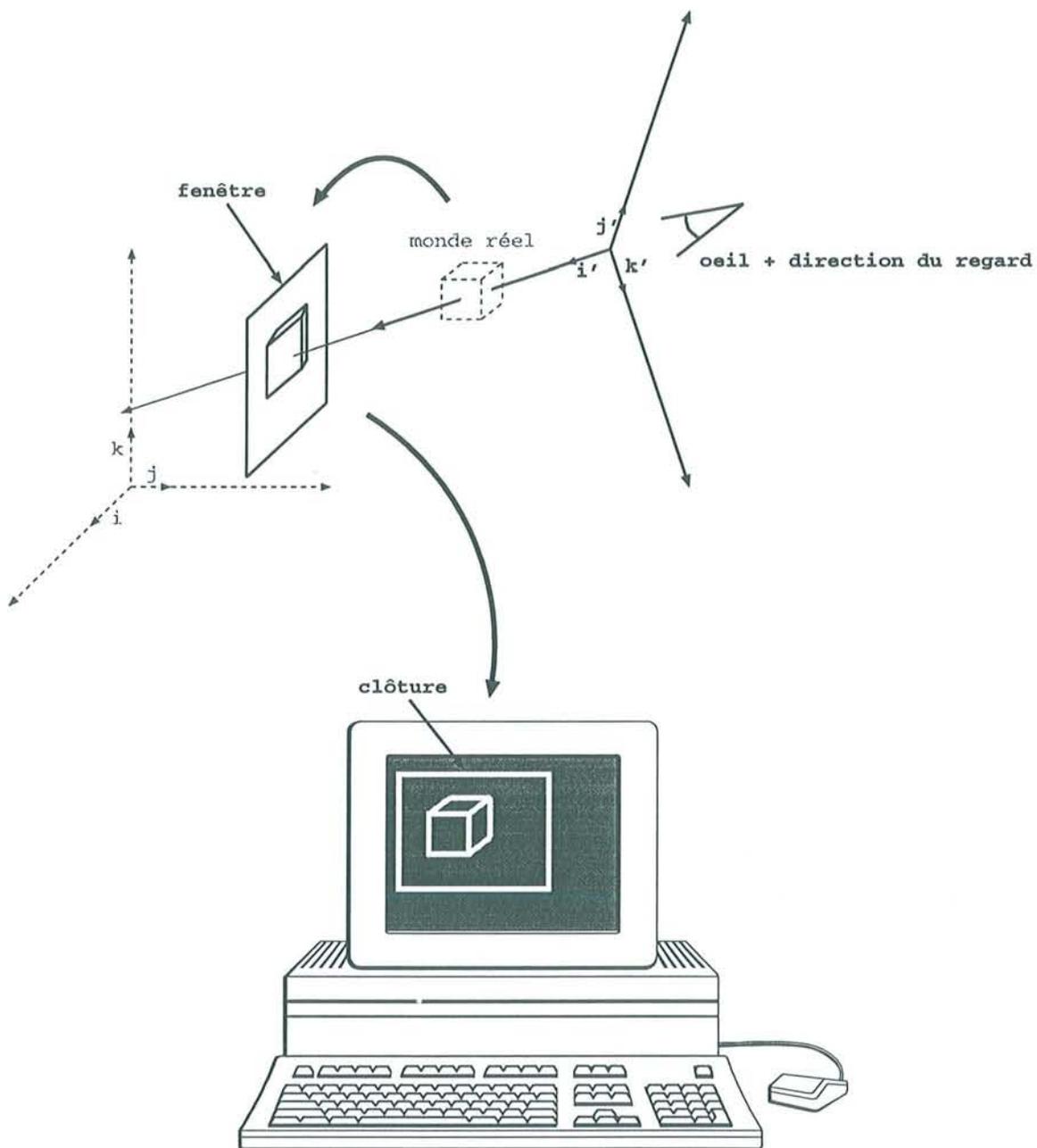


FIG. 5.2 – Les trois étapes de la visualisation.

d'effectuer les calculs sur les contours des solides de l'univers réel. Les entités peuvent être affichées suivant quatre modes :

- fil de fer (figure 5.3.(a)),
- fil de fer sans faces arrières (figure 5.3.(b)),
- facettes planes + éclairage constant + élimination des facettes arrières + tracé des arêtes (figure 5.3.(c)),
- rendu réaliste : les facettes sont éclairées par une source lumineuse placée à l'infini, les faces arrières sont éliminées (figure 5.3.(d)). Pour ce faire, le modèle de Lambert

a été utilisé⁶. Nous nous sommes limité à ce type de rendu, laissant de côté des modèles comme celui de Gouraud ou Phong⁷ qui auraient considérablement ralenti les performances de notre système.

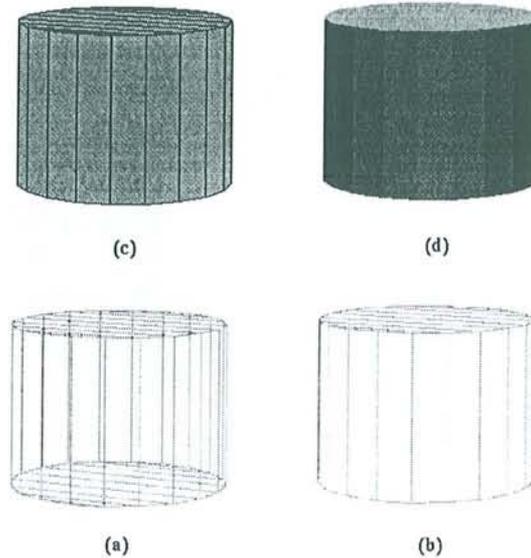


FIG. 5.3 – Les quatre modes graphiques possibles.

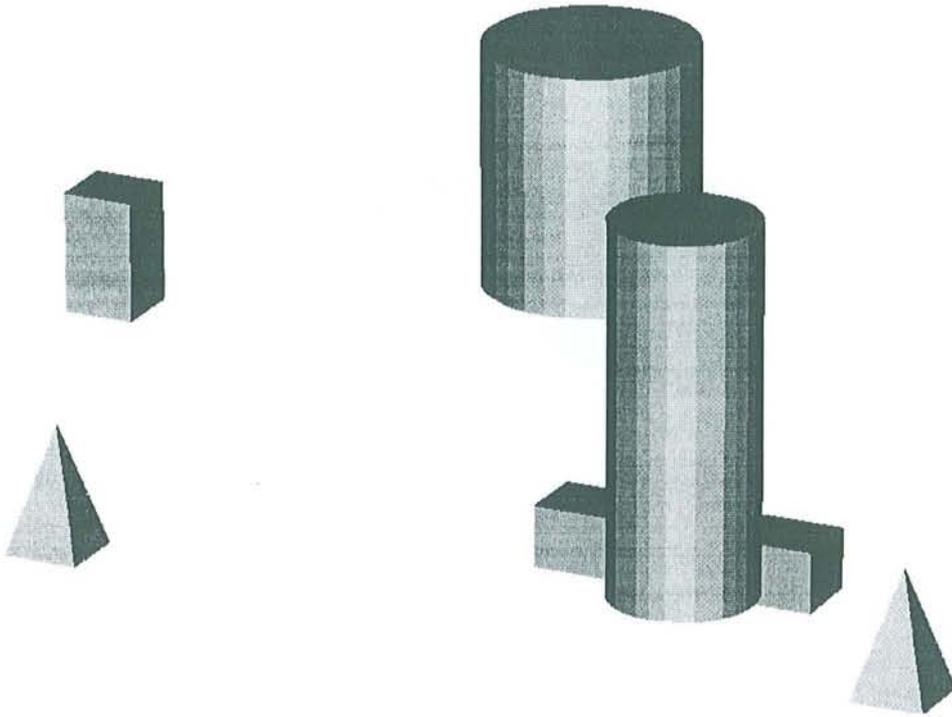
Les temps de calcul sont bien entendu croissants suivant le mode d'affichage utilisé, le rendu réaliste étant le plus gourmand (figure 5.4).

5.3.2 Bilan de ce modèle

Le système présenté est pratiquement indépendant de la plate-forme informatique sur laquelle il fonctionne. Par ailleurs, il semble facilement adaptable et évolutif. En ce qui concerne sa portabilité, il est nécessaire de préciser que l'ensemble du modèle est entièrement construit sur la base du tracé d'une droite dans une certaine couleur. Tout le reste des opérations, comme le remplissage d'un polygone par exemple est construit de manière incrémentale. Il suffit alors de changer la primitive de tracé de la droite pour pouvoir recompiler l'ensemble du système. Le système est écrit en Common-Lisp. Notre choix a porté sur ce dernier qui, même si il n'est pas aussi rapide que le langage C, présente de notre point de vue au moins deux avantages. Le premier procède d'une bonne évolutivité dans la perspective d'une prise en compte des aspects langagiers. Le second point concerne la maquette MultiDial qui a été développée dans notre équipe et que nous présenterons

6. Dans ce modèle, l'intensité de l'éclairage de chaque face est calculé par : $I_d = k * \frac{N \cdot L}{\|N\| \cdot \|L\|}$ où N est la normale à la face, L le vecteur directeur de la source de lumière, k une constante. Cette méthode suppose que la source de lumière se trouve à l'infini, (le produit scalaire $N \cdot L$ est alors constant sur toute la surface de la face).

7. Pour un exposé détaillé de ces modèles voir [Peroche *et al.*, 1990].

FIG. 5.4 – *Le mode de rendu réaliste.*

au paragraphe 5.4. Cette dernière est écrite pour une bonne part en Common-Lisp ce qui nous a aiguillé dans notre choix.

Par ailleurs, en ce qui concerne l'adaptabilité du système, le prototype développé semble évolutif. Ainsi, du côté purement tridimensionnel, le système est à même de prendre en compte des objets complexes non convexes ou troués. Les objets troués sont modélisés par l'orientation "matière à gauche" qui permettra de situer précisément les cavités d'un objet. Voir par exemple la figure 5.5.(b) qui représente une face plane qui comporte une cavité en son milieu. On crée alors une facette externe et une facette interne, cette dernière indiquant de par son orientation où se trouve la matière (en l'occurrence, à l'extérieur de la concavité). À titre indicatif, le système occuperait 2 mégas octets pour afficher environ 100 000 facettes comportant 4 côtés.

D'autre part, le modèle proposé semble adaptable à la simulation de nombre de systèmes faisant appel à un affichage graphique. À titre d'exemple, il suffit de considérer le cas d'un système de multifenêtrage qui permet d'afficher et de manipuler des fenêtres et icônes à l'écran. Étant donné que les fenêtres et les icônes peuvent se masquer entre elles partiellement ou totalement, ce type système relève d'une modélisation en $2D \frac{1}{2}$. On peut alors facilement simuler un tel système à l'aide de la maquette informatique proposée. Pour ce faire, les fenêtres et icônes sont construites dans le plan YoZ . Tandis qu'une icône est représentée par une facette plane, une fenêtre se trouve en revanche représentée par 2 à 4 facettes, une pour le corps de la fenêtre, une pour l'entête et éventuellement une à deux facettes pour l'ascenseur horizontal et vertical (figure 5.5.(c)). Les facettes associées à l'entête de la fenêtre et aux éventuels ascenseurs sont construites avec une abscisse plus

élevée que celle du corps de la fenêtre (figure 5.5.(c)). L'œil étant placé sur l'axe Ox , la visualisation de la scène donne l'impression de se retrouver en présence d'un système de multifenêtrage (figure 5.5.(a)).

Nous envisageons à court terme d'utiliser cette plate-forme pour évaluer l'algorithme de traitement des références proposé au précédent chapitre. Dans un second temps, nous intégrerons ce dernier au sein de la maquette MultiDial que nous allons présenter au prochain paragraphe.

Cette seconde phase d'implémentation se doublera de l'intégration de la plate-forme graphique au sein du module de dialogue de MultiDial, qui fournira à ce module une vision propre des objets de la tâche.

5.4 La maquette MultiDial

La maquette que nous allons décrire à présent est issue du projet ESPRIT MultiWorks (ESP 2105). L'ensemble de ce projet avait pour but de définir les concepts sous-jacents à l'élaboration d'une station de travail multimédia capable de traiter des dialogues multimodaux. Dans ce cadre, [Gaiffe et Romary, 1990] ont développé une maquette capable de gérer le dialogue entre le concepteur d'un document et l'éditeur hypertexte MultiCard. MultiCard est un logiciel qui permet de manipuler des documents texte, audio, vidéo et de naviguer au sein de tels documents. Son utilisation est cependant très complexe d'où l'utilité de lui adjoindre un module de dialogue qui rendrait sa manipulation plus aisée. Pour situer les apports du système développé par Gaiffe et Romary, il convient de préciser les cinq types d'objets qui sont manipulés au sein de MultiCard au moyen de nombreux menus déroulants :

- *nœud*: document de base du document hypermédia. Dans MultiCard, il y a une différence entre la structure du nœud qui gère les liens et les scripts et le contenu multimédia du nœud. Un nœud peut donc contenir du texte, des images, de la vidéo, des sons, des boutons,
- *ancree*: portion d'un nœud sensible à certaines actions. Par exemple, il peut s'agir d'un mot dans un texte, d'une région dans une image, etc. qui obéissent à des actions de l'utilisateur,
- *groupe*: ensemble de nœuds ou de groupes. Il y a une analogie très forte entre la notion de groupe et celle de répertoire sous Unix,
- *lien*: canal de communication entre deux objets. Différents messages peuvent être envoyés par ce biais. Les objets connectés par des liens peuvent être des ancres, des nœuds ou des groupes,
- *script*: programme décrivant le comportement d'un objet. Tout nœud, ancre ou groupe possède un script associé qui décrit le comportement de l'objet relativement aux différents messages qu'il peut recevoir.

La maquette MultiDial qui pilote MultiCard se situe donc dans une perspective de dialogue finalisé. Les requêtes de l'utilisateur au système de dialogue peuvent être de

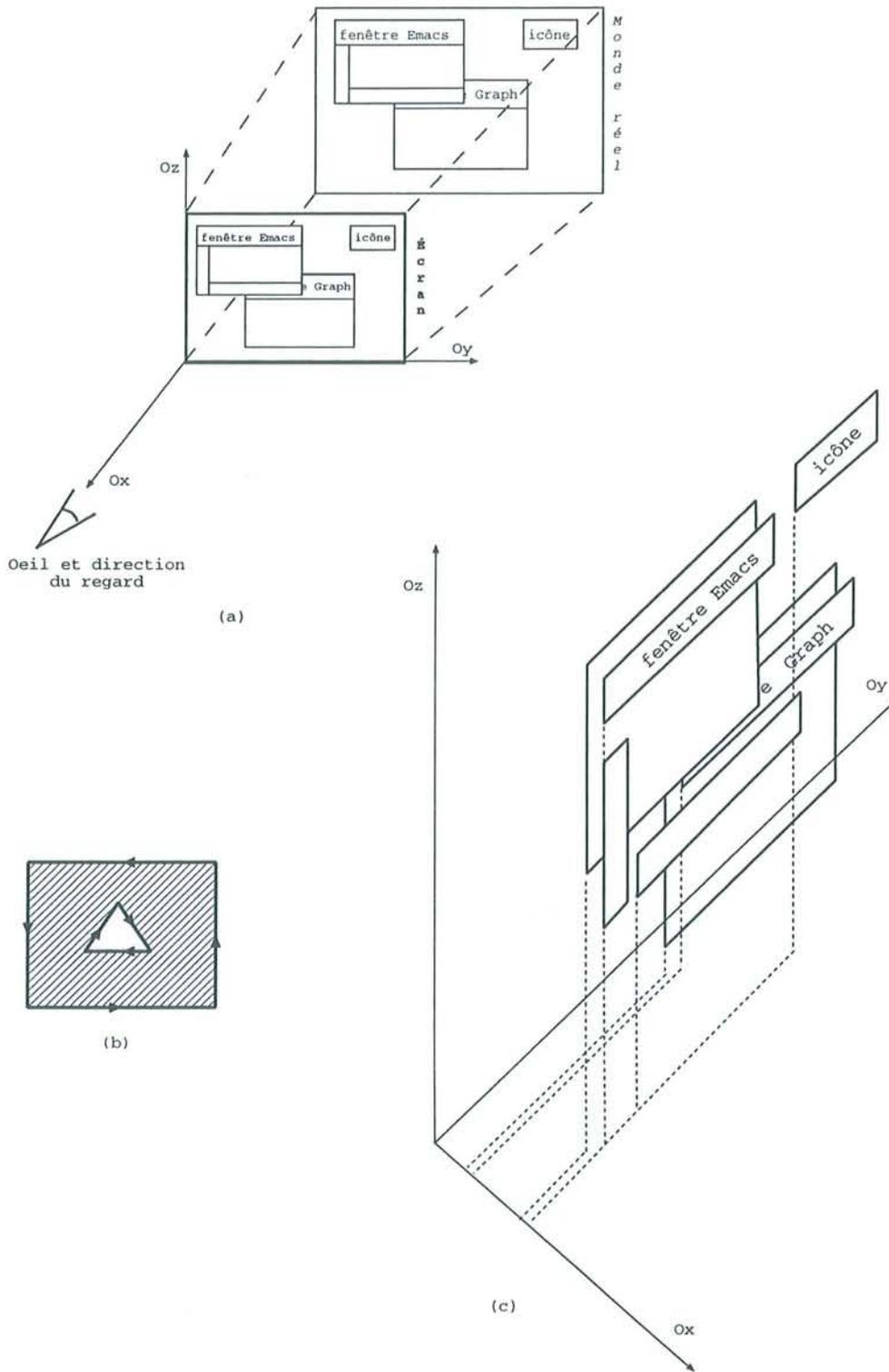


FIG. 5.5 – Adaptabilité du système de CAO.

deux types :

- dialogue de commande du type : *prédicat à l'impératif + arguments*,

- dialogue multimodal avec la possibilité de désigner des objets à la souris ou à l'aide d'un gant de désignation (ces deux aspects subissent actuellement une phase de mise à jour suivant les travaux de Bellalem déjà cités).

Le système de dialogue MultiDial est capable, pour l'instant, de créer/détruire des groupes d'objets, créer/détruire des nœuds et établir des liens entre ces objets (nœuds ou ancres). L'architecture du système est donnée à la figure 5.6. Le module MCEmacs permet d'éditer des fichiers textes. Le module MCEditor est adapté à l'édition de scripts attachés à un nœud ou une ancre. Raphaël est un module qui permet d'afficher des séquences audio-vidéo dans une fenêtre. Le traducteur est chargé de faire la liaison entre l'application (MultiCard) et le module de dialogue. Afin que le module de dialogue soit aussi indépendant que possible de l'application, un protocole a été mis en place qui repose sur l'envoi et la réception de messages. Ces messages peuvent par exemple permettre au module de dialogue d'interroger l'application MultiCard pour connaître l'ensemble des documents qui sont ouverts. En retour, MultiCard peut signaler qu'une opération s'est mal déroulée au module de dialogue qui agira en conséquence.

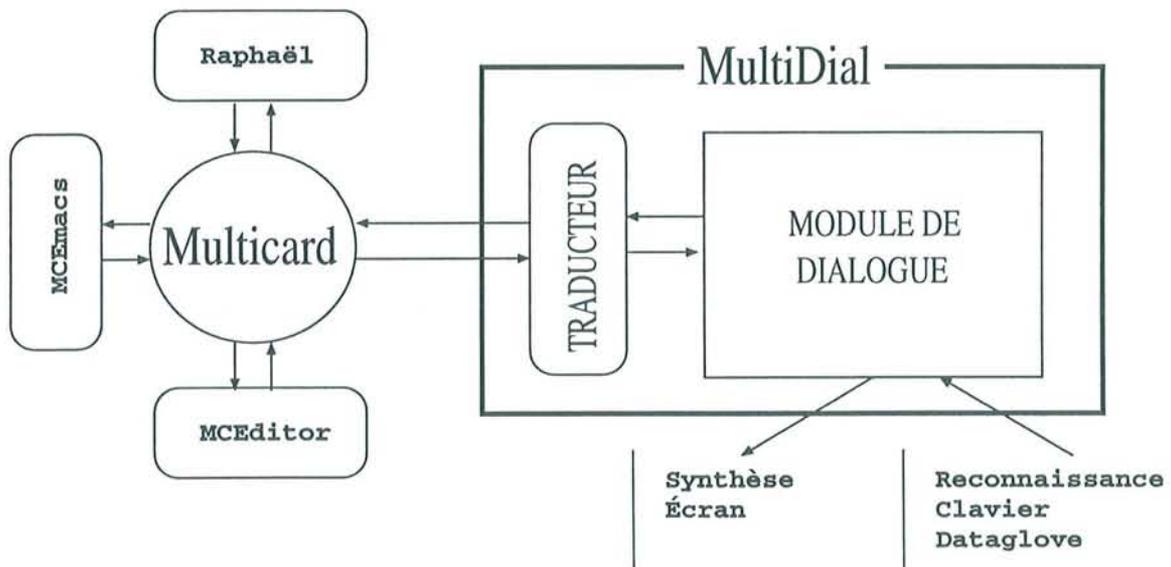


FIG. 5.6 – Architecture du système MultiDial.

Le module de dialogue de MultiDial fait appel à cinq sous-modules qui constituent une version simplifiée des connaissances et fonctions qui sous-tendent un système de dialogue tel que nous l'avons décrit au paragraphe 5.2 :

1. un module d'analyse syntaxique,
2. un modèle de la tâche,
3. un module de calcul des référents,
4. un historique du dialogue,
5. une représentation des éléments de la tâche. Cette représentation est propre au système de dialogue. Sa mise à jour se fait certes en corrélation avec l'état de l'ap-

plication mais indépendamment de celle-ci.

- Le module d'analyse syntaxique se fonde sur les grammaires d'unifications. Nous n'entrerons pas ici dans le détail des mécanismes sous-jacents aux grammaires d'unifications, le lecteur intéressé pourra se reporter à [Shieber, 1986].
- Le modèle de la tâche contient les pré-conditions attachées aux possibilités d'applicabilité de tel ou tel prédicat sur tel ou tel objet.
- Le module de calcul des référents se fonde sur les schémas d'axiologie définis par Gaiffe et Romary qui sont représentés à la figure 4.17 (page 150). Nous rappelons que dans ce type de formalisme, la recherche du(des) référent(s) s'effectue au sein de simples ensembles qui ne sont donc pas structurés, à la différence des cadres de référence.
- Les représentations de la tâche et de l'historique du dialogue reposent entièrement sur la notion de zone temporelle définie par Romary [1989] et [1991]. Une zone temporelle contient des faits et entretient des relations avec d'autres zones temporelles. Dans un tel modèle, seuls deux types de relations sont possibles : l'*inclusion* et l'*adjacence*. La relation d'inclusion est représentée à la figure 5.7.(a) et signifie que la zone Z2 est entièrement contenue dans la zone Z1. Autrement dit, Z2 n'a de sens que tant que Z1 existe. De la sorte, tous les faits valides dans Z1 sont valides dans Z2 ce qui peut se paraphraser par : "Z2 est incluse dans Z1 si et seulement si les informations contenues dans Z1 ont une durée plus longue que celles contenues dans Z2".

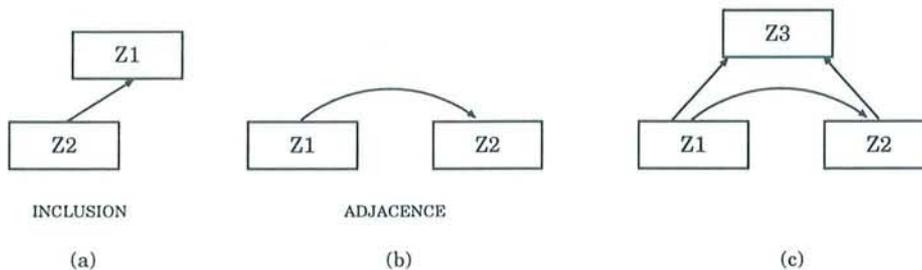


FIG. 5.7 – L'inclusion et l'adjacence temporelle.

L'adjacence qui est représentée par la figure 5.7.(b) indique que Z1 et Z2 sont temporellement disjointes et que Z1 précède Z2 temporellement. En particulier, les informations contenues dans Z1 peuvent être contradictoires avec celles contenues dans Z2. Il est alors impossible d'avoir une zone qui serait à la fois contenue dans Z1 et Z2. En revanche, Z1 et Z2 peuvent être à leur tour incluses dans une même zone Z3 selon le schéma de la figure 5.7.(c). Un tel schéma est très intéressant car il peut correspondre à deux états successifs d'un objet représenté par Z3. Si tel est le cas, cela correspond à la notion d'univers d'objet où la zone Z3 fait alors hériter les zones Z1 et Z2 d'un identificateur d'objet. Ainsi que nous le verrons ci-après, ce schéma permettra de prendre en compte les changements de propriétés associées à un objet.

Les propriétés d'un objet sont subordonnées au type ou à la catégorie de cet objet⁸. On obtient donc le schéma de la figure 5.8.(a) où #id désigne un nom de variable attaché à l'objet en vue de le distinguer d'autres objets ayant le même type et les mêmes propriétés. Ainsi, par exemple, dans le cadre d'un système de multifenêtrage, l'objet "fenêtre verte" sera représenté par la figure 5.8.(b).

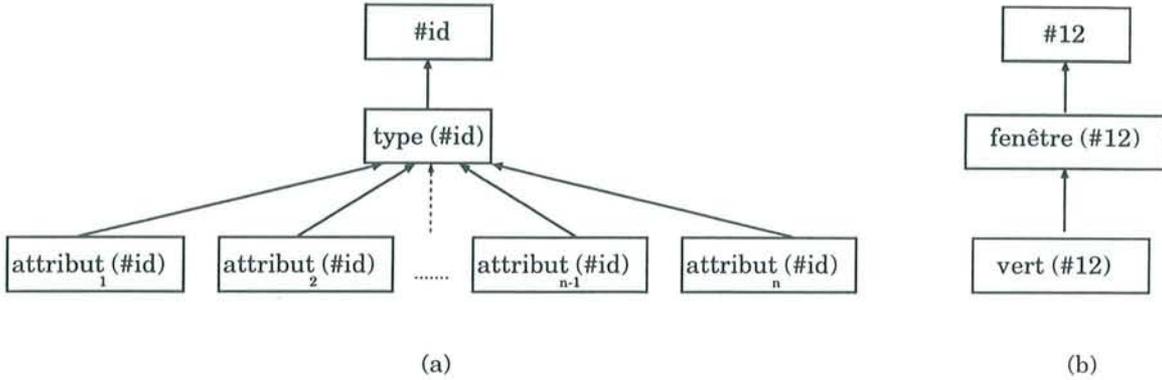


FIG. 5.8 – Représentation du type et des propriétés au sein des zones temporelles.

De la même manière, le changement d'une propriété d'un objet sera représenté suivant le schéma de la figure 5.9.(a). Ce type de schéma permet alors de coder une phrase du type "peints le carré vert en bleu" suivant la figure 5.9.(b).

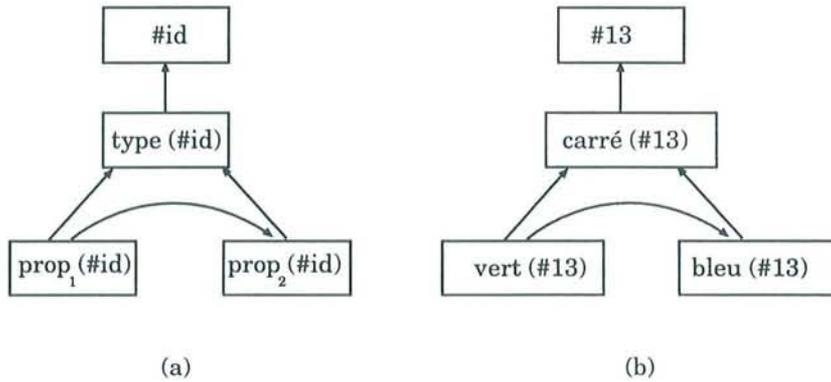


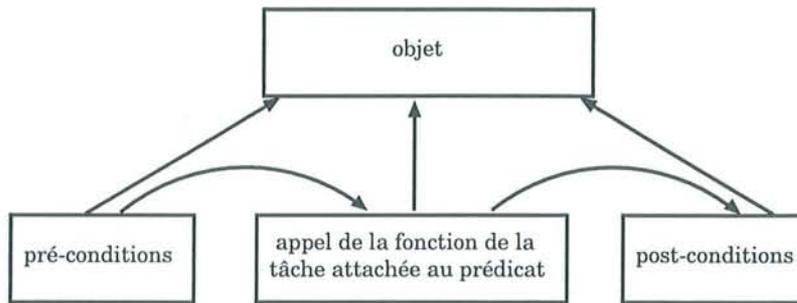
FIG. 5.9 – Représentation des changements au sein des zones temporelles.

Bien entendu, le seul état des objets n'est pas suffisant pour conserver l'évolution des objets de la tâche, il est en particulier également nécessaire de conserver les actions sur ces objets. Ce point nous amène à étudier comment sont pris en compte les prédicats au sein des zones temporelles.

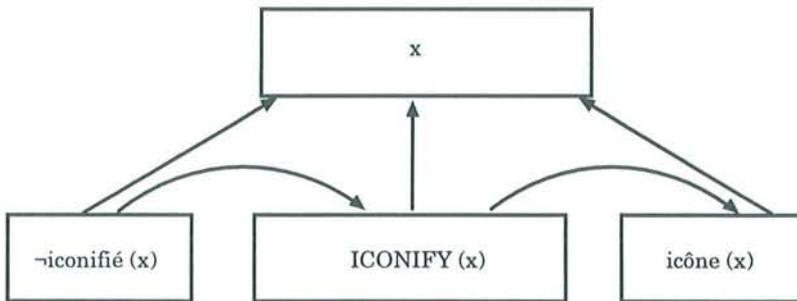
Tout changement, dans l'univers de la tâche, est relatif à un ou plusieurs objets dont on modifie les propriétés. Ces modifications peuvent porter sur les attributs

8. Dire d'un champ de maïs qu'il est vert n'est pas la même chose que de dire d'un homme qu'il est vert, à moins que ce dernier ne soit très malade ou maquillé pour un spectacle de clowns.

ou sur le type des objets comme dans la requête “*iconifie cette fenêtre*” où l’objet de type “*fenêtre*” change de type pour devenir une “*icône*”. Ces changements seront stockés au sein des zones temporelles sous la forme de post-actions. D’autre part, un prédicat ne s’applique pas nécessairement sur n’importe quel objet. Il est donc nécessaire de gérer certaines pré-conditions attachées à un prédicat⁹. On obtient finalement le schéma de la figure 5.10.(a) dont nous donnons un exemple à la figure 5.10.(b) pour la requête “*iconifie la fenêtre*”.



(a)



(b)

FIG. 5.10 – Les zones temporelles et le traitement des actions.

L’ensemble des schémas que nous avons décrit peut alors parfaitement se substituer à un historique de la tâche car ces schémas conservent l’état courant et passé des objets. Il ne reste plus alors qu’à ajouter à cet historique de la tâche, un historique du dialogue qui viendra pointer directement au sein des zones temporelles afin de conserver les différentes étapes du dialogue.

⁹. Les pré-conditions et post-conditions découlent du modèle de la tâche.

Bilan de ce modèle

Voici un petit exemple qui illustrera certaines des requêtes que la maquette MultiDial est capable de prendre en compte :

<i>Ouvre le nœud Mailbox.</i>	(174)
<i>Crées-y un bouton.</i>	(175)
<i>Ouvre le fichier Essai.</i>	(176)
<i>Mets le dans le script du bouton.</i>	(177)
<i>Ferme le nœud.</i>	(178)

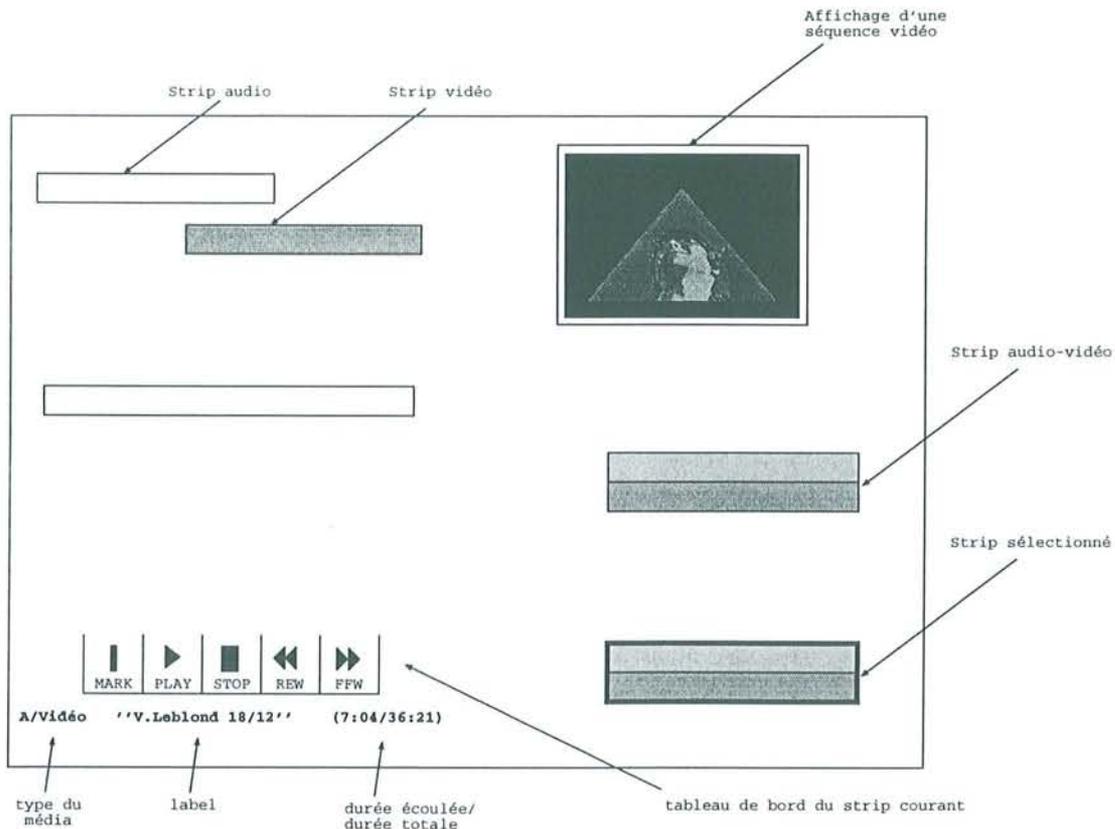
Deux aspects prédominants sont à retenir de ce modèle. Le premier aspect concerne la notion de zone temporelle qui offre une gestion simple des actions effectuées par l'application sur les objets manipulés au sein de l'application. Le second aspect procède de la séparation nette entre l'application et le module de dialogue. Pour gérer l'application, le module de dialogue se construit sa propre représentation des objets de la tâche. Cette séparation permet en théorie de créer un module de dialogue quel que soit l'application qui est pilotée. Pour ce faire, il suffit de changer le module de traduction ainsi que le modèle de la tâche qui contient en particulier les pré-conditions et les post-conditions attachées à chaque prédicat (il convient bien sûr d'adapter le lexique et éventuellement l'analyseur syntaxique).

La maquette MultiDial est actuellement en cours de modification afin de prendre en compte une nouvelle application de séquençement vidéo qui découle d'une collaboration entre le CRIN-CNRS-INRIA Lorraine et ALCATEL. Ce projet vise à élaborer une plate-forme multi-modale qui obéit à la parole ainsi qu'au geste et qui permet de réaliser différentes opérations sur des séquences audio, vidéo et audio-vidéo (appelées strips (voir la figure 5.11)). Ces opérations peuvent consister en des affichages de séquences vidéos, des duplications, des fusions (par exemple entre une séquence audio et une séquence vidéo), des séparations d'une séquence en plusieurs sous-séquences, etc. Les séquences sont représentées à l'écran sous forme de rectangles dont la longueur dépend de la durée de la dite séquence. L'utilisateur a également la possibilité de sélectionner un strip (cf. figure 5.11), s'il s'agit d'une séquence vidéo, le film s'affiche instantanément dans une fenêtre. La sélection d'un strip s'accompagne également de l'affichage de ses caractéristiques en bas de l'écran, l'utilisateur pouvant ensuite faire des avances rapides, marquer des événements, etc.

De notre point de vue, ce type de système se prête très bien à une multitude de références spatiales dont voici quelques exemples :

<i>Affiche le strip de droite.</i>	(179)
<i>Fusionne les strips de gauche.</i>	(180)
<i>Efface les séquences d'en bas.</i>	(181)

Par ailleurs, la représentation des strips vidéo sous forme de rectangles semble très bien se prêter à une gestion des références spatiales par notre méthode. Nous envisageons donc à court terme d'intégrer l'algorithme de calcul des référents spatiaux au sein de la

FIG. 5.11 – *Le séquenceur.*

maquette MultiDial tout en adaptant cette dernière à la nouvelle application de pilotage du séquenceur audio-vidéo. Par ailleurs, nous modifierons le modèle de représentation des objets de la tâche qui est propre au module de dialogue. La modification portera sur l'adjonction de la plate-forme graphique qui fournira au module de dialogue sa vision propre des objets de la tâche.

5.5 Conclusion

Après une brève présentation des composantes sous-jacentes à l'élaboration d'un système de dialogue, nous avons proposé une plate-forme graphique qui s'inspire d'aspects relatifs à la conception assistée par ordinateur. Cette plate-forme, que nous avons réalisée dans le cadre de cette thèse, est d'ores et déjà opérationnelle. Elle va nous permettre, dans un futur proche, de tester les algorithmes de calcul des référents spatiaux et ce, indépendamment de tout autre système informatique. Le choix de tel ou tel modèle de représentation des données est très important car l'un des points sensibles de l'intelligence artificielle réside dans le choix de représentations adéquates pour les connaissances manipulées. Ces représentations constituent non seulement le point de départ et le résultat des processus de calcul mais elles conditionnent également le type de déductions dont est capable le système informatique. Enfin, dans une dernière étape, nous avons présenté le

système MultiDial qui a été réalisé par Gaiffe et Romary. C'est au sein de ce système que nous allons intégrer notre méthodologie de recherche des référents spatiaux ainsi que notre plate-forme graphique qui fournira au modèle de dialogue sa propre vision des objets de la tâche.

Par ailleurs, on peut d'ores et déjà pressentir certaines difficultés lors de l'implantation informatique. La principale procède sans doute de la fusion de la notion de cadre avec celle de zone temporelle. La seconde concerne la prise en compte des aspects fonctionnels qui apparaîtront nécessairement au sein du pilotage de la plate-forme ALCATEL. Pour ces derniers, il semble nécessaire de mener une étude théorique préliminaire afin d'intégrer les aspects fonctionnels mis en jeu au sein de la notion de cadre de référence.

Conclusion et perspectives

« Celui qui en sait beaucoup sur les autres est peut-être instruit, mais celui qui se comprend lui-même est plus intelligent. Celui qui dirige les autres est peut-être puissant, mais celui qui s'est maîtrisé lui-même a encore plus de pouvoir ».

Tao-tö King.

Les travaux exposés dans cette thèse visent à apporter une contribution au domaine du dialogue homme-machine. Dans ce cadre, nous avons proposé une réflexion dont l'objectif est d'améliorer, au sein d'un système informatique, la compréhension des requêtes spatiales formulées par l'utilisateur. Parmi ces dernières, nous nous sommes focalisé sur celles qui faisaient intervenir des références spatiales purement langagières par opposition aux requêtes qui utilisent le langage naturel en le doublant d'un ou plusieurs geste(s) de désignation.

Cette étude nous a amené à dégager trois grands thèmes qui sous-tendent une partie de la réflexion proposée dans ce mémoire. Le premier thème concerne l'importance de la notion de dialogue au sein d'une interaction homme-machine, l'un des points essentiels concerne la conception que chaque partenaire entretient de l'autre. Ainsi, du côté de l'homme, la vision des capacités qu'il accorde à sa machine a une influence déterminante sur la manière dont il réalisera le(s) but(s) qu'il s'est fixé(s). Du côté de l'ordinateur, il faut cerner au mieux l'intention de l'utilisateur. Le deuxième thème procède de la notion de contexte qui possède un fort caractère pluridisciplinaire, ce qui la place au carrefour des sciences cognitives. Ainsi, pour le problème spécifique de la référence spatiale et plus particulièrement des prépositions spatiales, nous avons cherché à cerner précisément ce contexte à l'aide de la notion de cadre de référence. Enfin, le troisième thème concerne la notion de principe d'économie cognitive qui relève, de notre point de vue, d'une notion essentielle en intelligence artificielle. Outre son caractère pluridisciplinaire, cette notion est apte à subsumer des concepts tels que les maximes de Grice ou la notion de pertinence de Sperber et Wilson.

Une fois ces prémisses posées et afin d'isoler les principaux préceptes qui se dégagent de la modélisation de l'espace en langage naturel, nous avons examiné trois modèles qui relèvent respectivement d'une approche linguistique, logico-computationnelle et computationnelle. L'étude de ces modèles nous a permis d'insister sur l'importance des aspects fonctionnels ainsi que sur celle de cible et de site qui permettent de concevoir l'espace de manière plutôt relative qu'absolue.

Au vu de certains manques apparaissant dans ces trois modèles, nous avons cherché à proposer dans le troisième chapitre une autre méthodologie qui serait applicable à un système de représentation de figures géométriques en deux dimensions.

Le caractère très spécifique du modèle développé (en particulier, son incapacité à prendre en compte les évolutions discursives) nous a incité à élargir dans un premier temps notre réflexion au problème de la modélisation de l'espace en général pour proposer la notion de cadre. Dans un second temps, nous avons instancié cette dernière au problème de la référence spatiale ce qui nous a amené à développer la notion de cadre de référence pour laquelle nous avons proposé un algorithme qui autorise le calcul de certaines références spatiales et ce, au fil du discours.

Dans un dernier temps, nous avons présenté la maquette MultiDial au sein de laquelle sera implémentée la notion de cadre de référence ainsi que la maquette de conception assistée par ordinateur qui autorise une représentation souple des objets graphiques. Ce dernier chapitre s'ouvre sur un espace de recherche pour lequel nous envisageons d'ores et déjà différents objectifs à partir des travaux que nous avons présentés.

À court terme, nous devons tout d'abord mettre en œuvre totalement la notion de cadre de référence pour l'implanter au sein de la maquette MultiDial.

Nous envisageons également d'unifier nos travaux avec ceux de N. Bellalem qui portent sur l'introduction d'une composante gestuelle au sein du système MultiDial. La notion de cadre de référence semble en effet prometteuse pour modéliser certains aspects liés aux problèmes gestuels car une désignation sert essentiellement à attirer l'attention de l'interlocuteur sur un sous-espace du champ perceptif.

Sur un plan plus théorique, il nous faudra proposer une modélisation des prépositions spatiales telles que "*dans*" ou "*sur*". La prise en compte de ces prépositions passera nécessairement par un rôle accru des aspects fonctionnels au sein de la notion de cadre de référence. D'autre part, il nous faudra également mener en parallèle une réflexion sur la prise en compte du rôle du prédicat au sein des cadres de référence.

Il s'avère également nécessaire d'étudier plus en détail le rôle de la perception au sein de la méthodologie que nous avons proposée. Il semble en effet que certaines scènes graphiques très structurées puissent fortement influencer l'utilisateur lors de la formulation de sa requête. À titre d'illustration, il suffit de considérer une scène représentant un ensemble d'objets pratiquement alignés horizontalement, dans ce cas, une requête du

type : “*détruis le cinquième objet*” paraît tout à fait plausible. La prise en compte de telles requêtes nécessite alors d’appréhender précisément la scène affichée.

À plus long terme, nous souhaitons creuser davantage la notion de principe d’économie cognitive en la situant notamment plus finement par rapport à la notion de pertinence développée par Sperber et Wilson.

Enfin, une dernière piste de recherche concerne l’extension de notre modèle à la représentation d’aspects temporels. Les deux domaines de l’espace et du temps semblent en effet très liés, comme le montrent les expressions linguistiques par lesquelles ils sont exprimés. Il semble que ceci traduise une analogie plus profonde dont il faudrait tenir compte.

Annexe A

L'expérience de Milgram

Imaginez la situation suivante :

Recruté(e) par une petite annonce dans un journal local, vous vous rendez dans un laboratoire universitaire de psychologie pour prendre part à une expérience concernant l'apprentissage. On vous paie aussitôt les 150 francs que l'on vous avait promis pour votre participation. Un autre sujet, d'une cinquantaine d'années est arrivé avant vous pour prendre également part à l'expérience. L'expérimentateur explique que l'un de vous sera le moniteur et l'autre l'élève dans une recherche consacrée aux effets des punitions sur l'apprentissage de couples de mots. La punition sera un choc électrique. L'expérimentateur tire au sort et vous êtes désigné(e) comme moniteur. L'élève, quant à lui, se voit ligoté sur une chaise et des électrodes lui sont attachées au poignet. L'expérimentateur vous explique à tous les deux que le moniteur devra lire le premier terme d'une série de paires de mots et que l'élève devra essayer de donner le second terme correctement. Chaque fois que l'élève commettra une erreur, le moniteur devra lui donner une punition, c'est-à-dire un choc électrique.

On vous emmène personnellement dans une chambre et l'expérimentateur vous fait asseoir devant un générateur de chocs électriques. Celui-ci comprend une rangée de trente boutons allant de 15 à 450 volts : alors que les deux derniers boutons ont une étiquette "xx", les quatre précédents sont renseignés "danger, choc sévère". Les instructions sont de commencer au niveau le plus bas pour la première erreur et d'accroître l'intensité des chocs d'un bouton, c'est-à-dire de 15 volts à chaque erreur. L'élève commet tellement d'erreurs que, rapidement vous êtes amené à donner des chocs élevés. A partir de 75 volts, vous entendez l'élève grogner dans l'autre pièce chaque fois qu'il reçoit un choc. A 120 volts, il vous crie qu'il a eu réellement mal. A 150 volts, il rappelle, en hurlant, qu'il est sujet à des troubles cardiaques et que son coeur commence à le faire souffrir. Il supplie qu'on le laisse partir. A 270 volts, les hurlements font place à des cris d'agonie et ce jusqu'à 345 volts ; à partir de ce moment il n'y a plus de réponse audible. Suivant les instructions de l'expérimentateur, vous continuez à lire les mots et à traiter les silences de l'élève comme des erreurs, accroissant à chaque fois l'intensité des chocs de 15 volts jusqu'à atteindre l'échelon ultime de 450 volts.

Si durant l'expérience, vous hésitez à poursuivre la tâche et à délivrer les chocs, l'expéri-

mentateur vous dit selon le degré de résistance :

1. "Continuez, je vous en prie".
2. "L'expérience requiert que vous continuiez".
3. "Il est absolument essentiel que vous continuiez".
4. "Vous n'avez pas d'autre choix, vous devez continuer".

Si vous résistez encore après cette quatrième injonction, l'expérience est considérée comme terminée, de même qu'elle aurait pu l'être bien avant si vous aviez décidé d'abandonner. Auriez-vous été jusqu'au dernier choc? Stanley Milgram¹, l'auteur de cette expérience et d'une série d'autres études semblables, a posé la question à 39 psychiatres, 31 étudiants universitaires et 40 adultes de classe moyenne. *Aucun* ne déclara qu'il aurait été jusqu'au bout: un psychiatre et trois adultes avouèrent qu'ils auraient envoyé des chocs jusqu'à 300 volts, tandis que l'étudiant le plus obéissant prétendait qu'il n'aurait pas dépassé 210 volts. A la question, moins personnelle de prédire combien de gens en *général* iraient jusqu'à la fin, les réponses furent plus audacieuses: les psychiatres, des experts donc, prédirent *une* personne sur 1000. Ils fonctionnaient avec la théorie implicite de personnalité que l'homme est fondamentalement bon et raisonnable et, s'il se trouve quelqu'un pour aller jusqu'à 450 volts, il doit nécessairement s'agir d'un cas déviant, ayant une personnalité pathologique (ou appartenant à une nation non civilisée).

Quels furent les résultats? Milgram a conduit plusieurs expériences avec un scénario à peu près identique à celui que nous avons retracé et ce, avec des sujets masculins et féminins, de tous âges et de toutes professions. Le résultat qui ressort avec une régularité frappante est le suivant: 65 % des gens vont jusqu'à 450 volts lorsque le protocole d'expérience est semblable à celui qui vient d'être résumé. En d'autres termes, si l'on peut être plus clair, cela signifie que, sur simple injonction d'un expérimentateur, près de deux personnes sur trois acceptent d'adopter un comportement létal. N'oublions pas en effet, que l'élève avait des problèmes cardiaques et qu'à partir d'un certain moment il a cessé de répondre. Bien sûr, l'élève ne recevait pas de chocs électriques et ses cris relevaient d'une admirable mise en scène.

Notons enfin que les Méditerranéens réagissent comme les Américains et les Allemands. Des expériences menées en Italie, en Allemagne et aux U.S.A ont donné des résultats tout à fait semblables. D'autre part, comment invoquer massivement la personnalité des sujets devant les résultats suivants :

1. Lorsque c'est un chercheur scientifique qui reçoit les chocs électriques et supplie qu'on le laisse partir, plutôt qu'un quidam comme celui qui envoie les chocs, *personne* ne va jusqu'à 450 volts.
2. Si la victime n'est pas seulement audible, mais visible, qu'elle est assise à côté du moniteur qui doit notamment veiller à ce que les prétendues électrodes restent bien attachées, le taux d'obéissance descend de 65 % à 30 %. Selon une autre perspective, on pourrait dire que seulement deux personnes sur trois résistent à l'obéissance.

1. [Milgram, 1974] S. Milgram. *Soumission à l'autorité*. Paris: Calmann-Lévy, 1974.

3. Lorsque les chocs électriques restent ce qu'ils sont, qu'ils s'étalent donc de 15 volts à "danger, choc sévère" et puis à "xx, effet inconnu", mais que les réactions de la victime ne sont pas perceptibles, la *quasi-totalité* des sujets va jusqu'à "xx 450 volts".

En possession de ces résultats, comment incriminer la personnalité lorsque, selon les conditions expérimentales auxquelles sont confrontées les sujets, le taux d'obéissance varie de 0 à 100 %. Il devrait être clair que, quelle que soit notre théorie implicite de la personnalité, c'est la situation qui est spécialement en cause. D'ailleurs, parler seulement d'obéissance à l'autorité constituerait également une explication tronquée; cette obéissance varie selon les circonstances, plutôt qu'au gré de la personnalité.

Annexe B

La pensée latérale

Edward de Bono [1967] distingue la pensée verticale de la pensée latérale. La pensée verticale, qui relève de l'hémisphère gauche, traite l'information de manière spécifique. Elle procède analytiquement, réduisant le tout à ses éléments, elle est séquentielle, rationnelle, propositionnelle et temporelle. On la qualifie généralement de méthode hypothético-déductive.

La pensée latérale laisse tous les traits de la pensée verticale en suspens et joue avec les différentes *alternatives* possibles. Ses qualificatifs sont empruntés à ceux de l'hémisphère droit : a-temporel, diffus, visuo-spatial, intuitif et simultané.

Prenons un exemple tiré de Hampden-Turner [1990] pour illustrer le fonctionnement de la pensée latérale. Il y a de cela longtemps, lorsqu'on allait encore en prison pour dettes, un marchand se trouva de plus en plus endetté envers un usurier. L'usurier finit par avouer que son véritable but était d'obtenir la fille du marchand : il lui ferait grâce de sa dette si... En découvrant que le père et la fille étaient horrifiés par son offre, le rusé financier leur proposa de faire avec lui une promenade le long d'un chemin caillouteux et là, il suggéra que « *la Providence déciderait. Je prendrai deux pierres dans ce chemin, dit-il en se baissant prestement pendant qu'il parlait pour ramasser deux pierres, et je les mettrai (c'est ce qu'il fit), la blanche et la noire, dans ce sac. Tu n'iras pas en prison si ta fille, sans regarder, prend une pierre dans ce sac. Si c'est la pierre blanche, elle est libre et je te fais grâce de ta dette. Si c'est la noire, ta fille est à moi* ». À contrecœur, le marchand accepta, mais le cœur de sa fille se mit à battre follement. Elle était sûre, enfin presque, que le sac contenait deux pierres noires. Elle ne pouvait pas gagner l'épreuve. Démasquer le subterfuge de l'usurier équivalait à refuser de choisir et, donc, à envoyer son père en prison.

Le «choix» auquel la jeune avait à faire face était un exercice de «pensée verticale», une suite logique de «ou bien/ou bien», avec une catastrophe au bout, quel que fût le choix. Dans ce cas particulier, les prémisses sous-jacentes au «choix», avaient été combinées malhonnêtement, mais il arrive que les prémisses soient fausses même sans la contribution d'un scélérat ! Constamment nous faisons des erreurs dans les prémisses lorsque nous voulons résoudre des problèmes de toutes sortes. Du coup, la pensée verticale, qui est uni-directionnelle, est complètement à la merci de ses propres prémisses, justes ou fausses.

Voyons à présent comment la jeune fille du marchand qui est prise dans une double

contrainte a pu s'en sortir. Ce quelle a fait est un exercice de pensée latérale. Elle mit sa main dans le sac, prit un caillou et poussa un « *oh!* » en le faisant tomber accidentellement au milieu de centaines d'autres sur le chemin. « *Excusez-moi* », dit-elle, « *je suis si maladroite... mais vous pouvez connaître la couleur du caillou que j'ai choisi d'après la couleur de celui qui reste dans le sac* ».

En résumé, la pensée verticale converge vers les faits suivant une approche logico-méthodologique ; si les prémisses sont justes, elle fait gagner un temps considérable. La pensée latérale diverge ; elle considère les problèmes sous plusieurs angles, jusqu'à ce qu'une structure ou solution adéquate soit trouvée (si elle existe).

Annexe C

Élaboration des groupes homogènes perceptifs

Nous allons tout d'abord distinguer trois types de groupes homogènes perceptifs : les groupes homogènes perceptifs de type *droite/gauche* (figure C.1.(c)), les groupes homogènes perceptifs de type *haut/bas* (figure C.1.(d)) et les groupes homogènes perceptifs de proximité (figure C.1.(d)).

Chacun de ces groupes sert respectivement pour une préposition spatiale qui fait intervenir une relation de latéralité (exemples aux phrases (182) et (183)), une relation de verticalité ((184) et (185)) et enfin, tout autre relation spatiale qui ne fait pas intervenir la perspective¹ ((186) et (187)) :

Détruis les carrés de droite. (182)

Détruis les triangles qui se trouvent à gauche du rectangle. (183)

Agrandis les cercles d'en bas. (184)

Détruis les cercles qui se trouvent en dessous du cercle. (185)

Agrandis les cercles qui se trouvent en bas à droite. (186)

Détruis les triangles qui se trouvent en haut à gauche du cercle. (187)

C.1 Construction des groupes homogènes perceptifs de type *droite/gauche*

Nous rappelons à la figure C.2.(b) les difficultés auxquelles se heurte Briffault [1992] pour une requête du type "*détruis les carrés de droite*". La méthodologie de Briffault isole deux types de groupes homogènes perceptifs : le groupe $G1$ ou les groupes $G1'$, $G2'$ et $G3'$ selon : $G1 = \{C1, C2, C3, C4\}$, $G1' = \{C1\}$, $G2' = \{C1, C2\}$ et $G3' = \{C4\}$. Tandis que

1. Nous rappelons que le chapitre 3 ne se préoccupe pas des relations du type "*devant*" / "*derrière*".

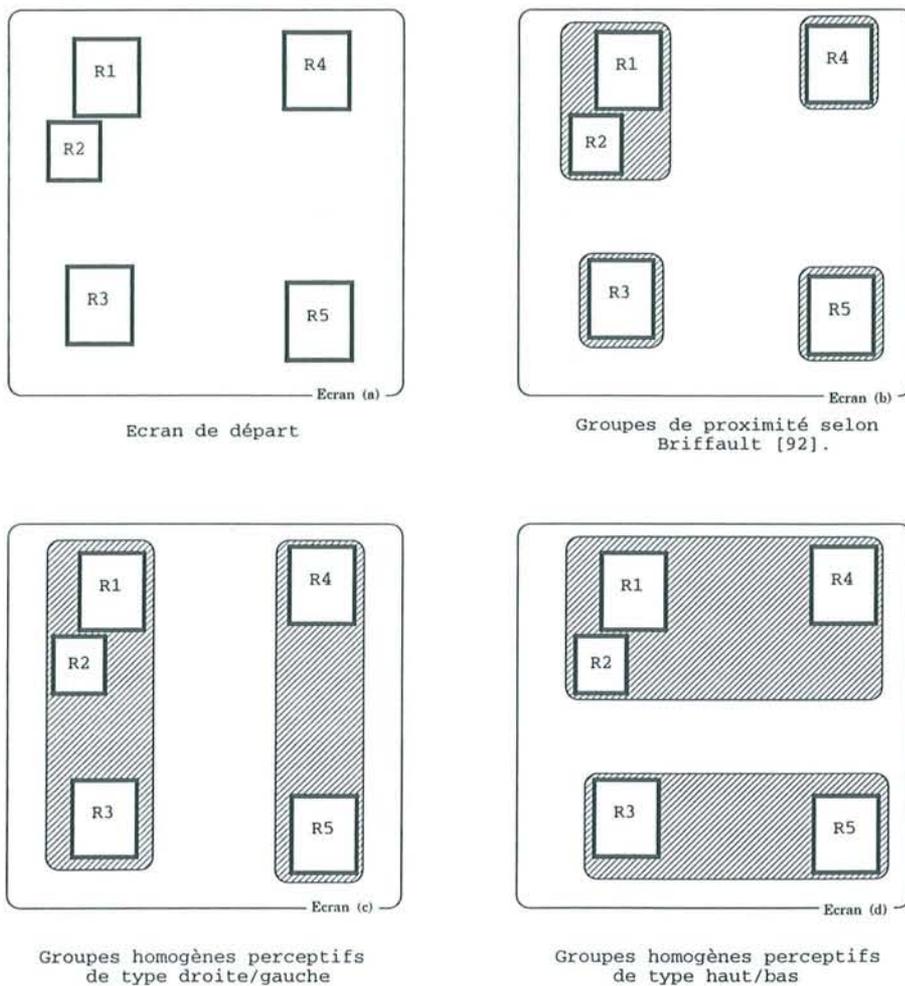


FIG. C.1 – *Groupes de perception homogènes.*

le premier groupement est trop grossier, le second est trop fin car il apparaît opportun de rassembler les carrés C2, C3 et C4 au sein du même groupe homogène perceptif.

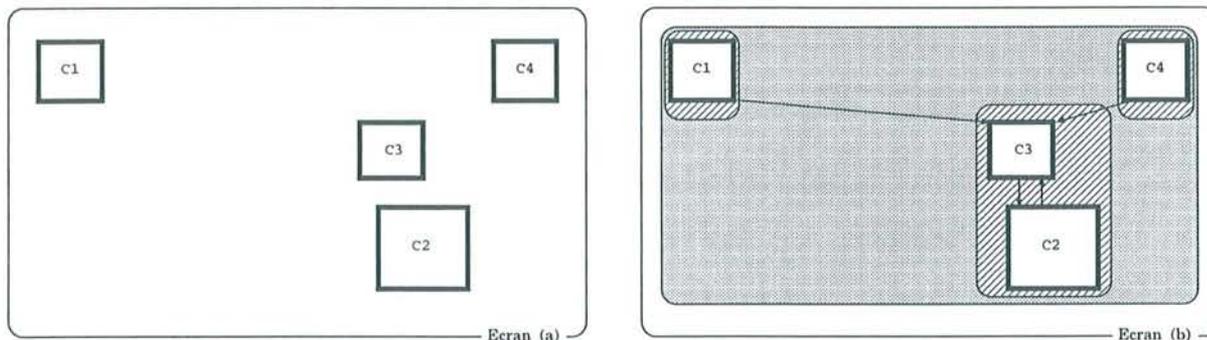


FIG. C.2 – *Comment construire les groupes homogènes perceptifs?*

Au vu de ce constat, une solution possible consisterait à relier deux objets A et B par la relation $Pp_{D/G}$ (plus proche droite/gauche) si la distance qui les sépare ne dépasse pas

un certain seuil. Il est important de noter que la distance $Pp_{D/G}$ est une relation qui opère sur la base de la distance horizontale qui sépare deux objets, faute de quoi, les groupes homogènes perceptifs ne seraient pas construits correctement. Il suffit de considérer la figure C.3 où la distance euclidienne classique ne permettrait pas de rassembler les carrés C1 et C3 dans le même groupe homogène perceptif pour une requête du type “*détruis les carrés de droite*” car la distance séparant C1 de C3 est plus grande que celle séparant C2 et C3.

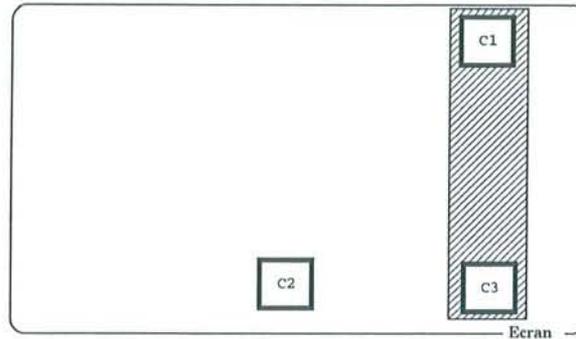


FIG. C.3 – La relation $Pp_{D/G}$.

Le fait d'imposer un seuil permet alors de ne pas relier le carré C1 au carré C3 à la figure C.2.(b). De manière plus générale, la difficulté réside dans la valeur à affecter au seuil. En effet, on peut toujours trouver un contexte particulier qui mettra en défaut cette valeur. Il suffit de considérer la figure C.4 pour une requête du type “*détruis les carrés de droite*” où une distance de 1/5 de la largeur de l'écran semble satisfaisante à la figure C.4.(a) alors qu'elle ne l'est plus à la figure C.4.(b).

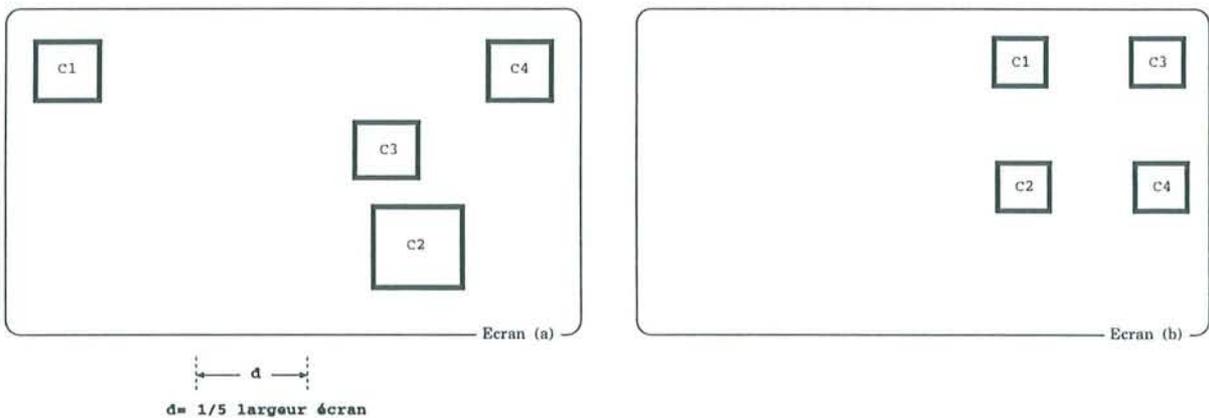


FIG. C.4 – Le problème de la granularité au sein des groupes homogènes perceptifs.

Il paraît cependant évident qu'il existe une distance entre les objets en deçà de laquelle l'utilisateur n'effectue plus de structuration perceptive (il range alors les objets concernés au sein du même groupe homogène perceptif). Nous proposons d'assigner la valeur 1/5 au seuil, les tests empiriques que nous avons pu mener nous indiquent qu'il s'agit d'une bonne valeur de départ. Par ailleurs, pour les cas qui posent problème, tel que celui qui

a été mis en évidence à la figure C.4.(b), nous proposons de diminuer la valeur du seuil jusqu'à isoler au moins deux groupes homogènes perceptifs.

C.2 Construction des groupes homogènes perceptifs de type *haut/bas*

Il suffit d'utiliser la relation $Pp_{H/B}$ (plus proche haut/bas) qui effectue ses calculs à partir de la distance verticale séparant deux objets. Enfin, le facteur $1/5$ est considéré relativement à la hauteur de l'écran (ce facteur peut bien sûr diminuer lorsqu'un seul groupe homogène perceptif a été isolé).

C.3 Construction des groupes homogènes perceptifs de proximité

On utilise la relation Pp proposée par Briffault [1992] en ne regroupant que des objets qui sont séparés d'une distance inférieure à $1/5$ de la longueur de la diagonale de l'écran (de la même manière que précédemment, si on ne parvient qu'à isoler un seul groupe homogène perceptif, il faut diminuer cette valeur).

Nous menons actuellement une réflexion en vue d'améliorer les algorithmes proposés. Deux types de travaux nous semblent applicables. Il s'agit tout d'abord des diagrammes de Voronoi² [Preparata et Shamos, 1985] qui ont été récemment étendus afin de prendre en compte des objets segmentés et plus seulement des nuages de points. D'autre part, les méthodes de classification des données de type ACP (analyse en composantes principales) semblent également prometteuses (voir en particulier [Celeux *et al.*, 1989]).

2. L'ensemble de tous les points plus proches d'un site donné que de tous les autres sites d'un ensemble est une structure géométrique importante appelée le *polygone de Voronoi* de ce site. L'union de tous les polygones de Voronoi d'un nuage de sites est appelé le *diagramme de Voronoi*.

Table des figures

1.1	Prépositions spatiales et contexte.	30
1.2	Visualisation d'un salon pour une application d'aménagement.	36
1.3	Une classification possible des commandes multimodales vocales et gestuelles.	38
1.4	Exemple de deux labyrinthes pour deux joueurs A et B.	42
2.1	Un "dans" plus insaisissable qu'il n'y paraît.	48
2.2	La notion de ressemblance de famille.	51
2.3	Les trois directions de base dans le repère du corps humain.	54
2.4	Les orientations en miroir et en tandem.	55
2.5	Deux illustrations de la préposition "derrière".	55
2.6	* La lampe est sur la table.	58
2.7	* Le balai est sur le mur.	58
2.8	Changements de classe des verbes de mouvement.	60
2.9	Les sept zones de localisation d'après Asher et Sablayrolles 1995.	61
2.10	Relations topologiques.	65
2.11	Le modèle de Hernandez.	66
2.12	Les secteurs d'acceptabilité.	67
2.13	Zones d'acceptabilité non homogènes.	68
2.14	Prise en compte des différentes orientations dans le modèle de Hernández.	69
3.1	Le problème de la référence aux actions.	77
3.2	"le N".	83
3.3	De l'intérêt de l'axiologie " <i>fenêtre</i> " versus \neg " <i>fenêtre</i> ".	84
3.4	"ce N".	85
3.5	"un N".	85
3.6	Schémas d'axiologie.	87
3.7	Le problème des propriétés non intrinsèques.	87
3.8	Avantages d'une optique différentielle.	89
3.9	"Actes gestuels" <i>versus</i> "actes langagiers".	90
3.10	Quand le concept de cible empiète sur celui de site.	93
3.11	Unification de la référence définie et indéfinie.	95
3.12	Un modèle imprécis.	98
3.13	Acceptabilité des prépositions selon Briffault.	99
3.14	Différentes acceptabilités de "à gauche de".	100
3.15	Deux nuages de probabilité pour la préposition "à gauche de".	101

3.16	Imprécision des centres de gravité.	101
3.17	L'approche de Wazinski, ses limites.	102
3.18	Les problèmes de la référence absolue et plurielle.	104
3.19	Le problème de la portée localisatrice limitée du site.	105
3.20	De la difficulté à construire le prototype.	107
3.21	Les errements du prototype.	108
3.22	Le cas de la référence à plusieurs objets.	109
3.23	Modèle de calcul des référents "absolu" pour la préposition " <i>de droite</i> ".	110
3.24	Ambiguïté possible pour les références indéfinies.	111
3.25	Modèle de calcul des référents "absolu" pour la préposition " <i>en haut à droite</i> ".	111
3.26	Modèle de calcul des référents "relatif" pour la préposition " <i>à droite de</i> ".	112
3.27	Modèle de calcul des référents "relatif" pour la préposition " <i>en haut à droite de</i> ".	114
3.28	Modèle de calcul des référents pour une localisation à partir d'un groupe d'entités.	115
3.29	Construction des groupes homogènes perceptifs selon Briffault.	116
3.30	Structuration spontanée d'un ensemble d'éléments (d'après [Delay et Pichot, 1984]).	117
3.31	Comment construire les groupes homogènes perceptifs?	117
3.32	Groupes de perception homogènes.	118
3.33	Apports du modèle développé.	120
3.34	Ambiguïté dans la détermination du groupe localisateur.	121
3.35	De la nécessité des cadres.	122
4.1	Limitations de la portée de la préposition " <i>devant</i> ".	129
4.2	Nécessité des cadres dans un système de multifenêtrage.	130
4.3	De la nécessité des cadres.	130
4.4	De la pertinence des cadres de référence.	134
4.5	Différents points de vue sur une île.	135
4.6	De l'intérêt des cadres de référence.	137
4.7	Deux directions pertinentes: la latéralité et la verticalité.	137
4.8	La rupture au sein des cadres de référence.	139
4.9	Raisonnement en priorité dans le précédent cadre de référence.	139
4.10	Pertinence des cadres de référence.	141
4.11	Deux illustrations de la préposition " <i>au-dessus de</i> ".	142
4.12	Construction du "search domain" en considérant les objets environnants.	144
4.13	Les évolutions discursives des univers spatiaux.	145
4.14	Les évolutions discursives des univers spatiaux.	146
4.15	Connaissances nécessaires à l'interprétation d'une référence spatiale.	148
4.16	Inclusion partielle de la cible dans le focus de recherche.	149
4.17	Les trois schémas d'axiologie selon Gaiffe [92].	150
4.18	Le cas de la rupture au sein des cadres de référence.	152
4.19	Le problème de la portée localisatrice limitée du site.	153
4.20	Exemple de site non obstruant.	154
4.21	Limitation de la portée localisatrice d'un site par un autre.	154

4.22	Focus et extension associés à la préposition “ <i>de droite</i> ” en 2D et en 3D.	155
4.23	Différents usages de la préposition “ <i>sur</i> ”.	156
4.24	La préposition “ <i>sur</i> ” et la relation de contact.	156
4.25	Image mentale du focus et extension associés de la préposition “ <i>à droite de</i> ”.	157
4.26	Différentes orientations intrinsèques.	158
4.27	Construction du cadre prototypique associé à un ou plusieurs sites.	161
4.28	Le problème des enveloppes englobantes.	162
4.29	Une application manipulant des figures géométriques.	164
4.30	Instanciación du nouveau cadre de référence et affinage.	167
4.31	Représentation du champ “prédiction des alternatives” défini par $\{C\} \cup \neg\{C\}$ à l’aide de la notion d’axiologie.	167
4.32	Une levée d’ambiguïté sur la base du champ “prédiction des alternatives”.	168
4.33	Une scène représentant des figures géométriques.	168
4.34	Détermination du champ “prédiction des alternatives”.	170
4.35	Image perçue comme étant composée de colonnes, la similitude des brillances des points l’emporte sur la proximité.	174
4.36	Un système de multifenêtrage.	177
4.37	Algorithme de calcul des références spatiales.	182
4.38	Une scène d’objets.	183
4.39	Exemple de fonctionnement des cadres de référence.	185
4.40	Exemple de fonctionnement des cadres de référence.	188
4.41	Exemple de fonctionnement des cadres de référence.	190
4.42	Modèle abstrait pour le positionnement.	196
4.43	Nuance entre le positionnement et la référence “pure”.	196
4.44	Cadres de référence et ambiguïté.	201
4.45	Un cas d’ambiguïté?	201
4.46	Visualisation d’un salon pour une application d’aménagement.	204
5.1	La convention “matière à gauche”.	211
5.2	Les trois étapes de la visualisation.	212
5.3	Les quatre modes graphiques possibles.	213
5.4	Le mode de rendu réaliste.	214
5.5	Adaptabilité du système de CAO.	216
5.6	Architecture du système MultiDial.	217
5.7	L’inclusion et l’adjacence temporelle.	218
5.8	Représentation du type et des propriétés au sein des zones temporelles.	219
5.9	Représentation des changements au sein des zones temporelles.	219
5.10	Les zones temporelles et le traitement des actions.	220
5.11	Le séquenceur.	222
C.1	Groupes de perception homogènes.	236
C.2	Comment construire les groupes homogènes perceptifs?	236
C.3	La relation $Pp_{D/G}$	237
C.4	Le problème de la granularité au sein des groupes homogènes perceptifs.	237

Bibliographie

- [Allen, 1983] J. F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.
- [Allen, 1984] J. F. Allen. Towards a general theory of action and time. *AI Journal*, pages 123–154, 1984.
- [Amalberti *et al.*, 1993] M. Amalberti, N. Carbonell, et P. Falzon. User representations of computer systems in human-computer interaction. *Man-machine Studies*, (38):547–565, 1993.
- [Andler, 1992] D. Andler. *Introduction aux sciences cognitives*. Folio, 1992.
- [Andrès, 1993] M. Andrès. Une représentation des figures et de leurs déplacements dans Compèrobot. Dans *Les cahiers du LAIAC*, volume 13, pages 1–14. 1993. Caen - France.
- [Arnold et Lebrun, 1992] M. Arnold et C. Lebrun. Utilisation d’une langue pour la création de scènes architecturales en image de synthèse, expérience et réflexions. Dans *Intellectica*, volume 99, pages 151–186. 1992.
- [Asher et Sablayrolles, 1995] N. Asher et P. Sablayrolles. A typology and discourse semantics for motion verbs and spatial prepositions in french. *Semantics*, 12(2):163–209, 1995.
- [Aurnague *et al.*, 1993] M. Aurnague, L. Vieu, et A. Borillo. Vers une représentation formelle des concepts spatiaux dans la langue. Dans *Images et Langages. Multimodalité et Modélisation Cognitive*, pages 109–120. Colloque interdisciplinaire du CNRS, Paris, 1er et 2 Avril 1993.
- [Aurnague et Vieu, 1993] M. Aurnague et L. Vieu. A three-level approach to the semantics of space. Dans *The semantics of prepositions*, pages 393–439. Mouton de Gruyter, 1993. Cornelia Zelinsky-Wibbelt (Editor).
- [Aurnague, 1991] M. Aurnague. *Contribution à l’étude de la sémantique formelle de l’espace et du raisonnement spatial : la localisation interne en français, sémantique et structures inférentielles*. 1991. Thèse d’Université Paul Sabatier Toulouse.
- [Austin, 1970] J. L. Austin. *Quand dire, c’est faire*. Seuil, 1970.

- [Beaugé, 1995] L. Beaugé. *Définition de mécanismes de mémorisation pour les systèmes neuromimétiques, inspirations pluridisciplinaires*. 1995. Thèse d'Université Henri Poincaré - Nancy I.
- [Bellalem *et al.*, 1995] N. Bellalem, L. Romary, et D. Schang. Which representation for a proper treatment of referring expressions in a man-machine multimodal dialogue? *ESCA Workshop on Spoken Dialogue Systems, Theories and Applications, Denmark*, pages 61-64, 1995.
- [Bellalem et Romary, 1993] N. Bellalem et L. Romary. Le dialogue homme-machine multimodal : vers la compréhension du geste de désignation. *L'Interface des mondes réels et virtuels, 2^{ème} Conférence Internationale, Montpellier*, pages 217-227, 22-26 Mars 1993.
- [Bellalem, 1995] N. Bellalem. *Etude du mode de désignation dans un dialogue homme-machine à forte composante langagière : analyse structurelle et interprétation*. 1995. Thèse d'Université Henri Poincaré - Nancy I.
- [Berne, 1990] E. Berne. *Analyse transactionnelle et psychothérapie*. Payot, 1990.
- [Bierwish, 1983] M. Bierwish. Semantische und konzeptuelle repräsentation lexicalischer einheiten. Dans *Untersuchungen zur Semantik*, pages 15-59, 1983.
- [Boers, 1987] F. Boers. Behind, beyond, under, underneath, beneath, below: a descriptive and explanatory study of spatial and non-spatial senses. Dans *Antwerp papers in linguistics*, volume 53. 1987.
- [Bourguet, 1992] M. L. Bourguet. IcpPlan : Dialogue multimodal pour la conception de plans architecturaux. pages 369-374, 1992. 19^{èmes} JEP Bruxelles.
- [Bouyer, 1997] E. Bouyer. *Quel lexique pour résoudre les références dans un dialogue homme-machine ?* 1997. Thèse d'Université Henri Poincaré - Nancy I.
- [Briffault, 1992] X. Briffault. *Modélisation informatique de l'expression de la localisation en langage naturel*. 1992. Thèse d'Université de Paris VI.
- [Briffault, 1993] X. Briffault. Modélisation informatique de l'expression de la localisation spatiale en langage naturel. Dans *Images et Langages. Multimodalité et Modélisation Cognitive*. Colloque interdisciplinaire du CNRS, Paris, 1er et 2 Avril 1993.
- [Bruce et Green, 1993] V. Bruce et P. Green. *La perception visuelle : physiologie, psychologie et écologie*. Presses Universitaires de Grenoble, 1993.
- [Carbonell *et al.*, 1985] N. Carbonell, J.P. Haton, J.M. Pierrel, et al. Les connaissances nécessaires dans un système de dialogue oral homme-machine. *5^{ème} congrès RF-IA, Grenoble*, 1985.
- [Carlson-Radvansky et Irwin, 1993] L. A. Carlson-Radvansky et D.E. Irwin. Frames of reference in vision and language: Where is above? Dans *Cognition*, volume 46, pages 223-244. 1993.

- [Carré *et al.*, 1991] R. Carré, J.F. Dégremont, M. Gross, J.M. Pierrel, et G. Sabah. *Langage Humain et Machine*. Presses du CNRS, 1991.
- [Celeux *et al.*, 1989] G. Celeux, E. Diday, G. Govaert, Y. Lechevallier, et H. Ralambondrainy. *Classification automatique des donnés*. Dunod, 1989.
- [Chapelier *et al.*, 1994] L. Chapelier, C. Fay-Varnier, A. Roussanaly, et V. Saint-Dizier. Recueil et analyse d'un corpus d'interactions multimodales. *ERGO-IA*, 1994.
- [Chapelier *et al.*, 1995] L. Chapelier, C. Fay-Varnier, et A. Roussanaly. Modeling an intelligent help system from a wizard of Oz experiment. *ESCA Workshop on Spoken Dialogue Systems, Theories and Applications, Denmark*, pages 101–104, 1995.
- [Chapelier, 1996] L. Chapelier. *Dialogue d'assistance dans une interface homme-machine multimodale*. 1996. Thèse d'Université Henri Poincaré - Nancy I.
- [Charolles, 1996] M. Charolles. Les univers spatiaux et leur extension. Dans (*Manuscrit provisoire*), 1996.
- [Corblin, 1987] F. Corblin. *Indéfini, défini et démonstratif*. Librairie DROZ, Genève-Paris, 1987.
- [Corblin, 1995] F. Corblin. *Les formes de reprise dans le discours, anaphores et chaînes de références*. Presses Universitaires de Rennes, 1995.
- [Cudicio, 1986] C. Cudicio. *Comprendre la PNL. La programmation neurolinguistique, outil de communication*. Les éditions d'Organisation, 1986.
- [Dauchy *et al.*, 1993] P. Dauchy, C. Mignot, et C. Valot. Joint speech and gesture analysis: some experimental results on multimodal interface. *EUROSPEECH, Berlin*, pages 1315–1318, 1993.
- [de Bono, 1967] E. de Bono. *The use of lateral thinking*. Harmondsworth, Penguin, 1967.
- [Delay et Pichot, 1984] J. Delay et P. Pichot. *Abrégé de psychologie*. Masson, 1984.
- [Denis, 1989] M. Denis. *Image et Cognition*. Presses Universitaires de France, 1989.
- [Dony, 1985] R. Dony. *Graphisme Scientifique sur micro-ordinateur de la 2ème à la 3ème dimension*. Masson, 1985.
- [Dowty *et al.*, 1981] D. R. Dowty, R.E. Wall, et S. Peter. *Introduction to Montague semantics*. Dordrecht : D. Reidel, 1981.
- [Dubois, 1991a] D. Dubois. Catégorisation et cognition : “10 ans après”, une évaluation des concepts de Rosch. Dans *Sémantique et cognition*, pages 31–54. Editions du CNRS, 1991.
- [Dubois, 1991b] D. Dubois. *Sémantique et cognition, Catégories, Prototypes et Typicalité*. Editions du CNRS, 1991.

- [Duermael, 1994] F. Duermael. *Référence aux actions dans des dialogues de commande homme-machine*. 1994. Thèse de l'Institut National Polytechnique de Lorraine - Nancy.
- [Falzon, 1986] P. Falzon. *Langages opératifs et compréhension opérative*. 1986. Thèse d'état, Université René Descartes, Paris.
- [Faure et Arnold, 1993] C. Faure et M. Arnold. L'interaction homme-machine du point de vue des principes d'économie. *IHM-93*, 1993.
- [Fodor, 1986] J. A. Fodor. *La modularité de l'esprit*. Les Editions de Minuit, 1986.
- [Gaiffe et al., 1991] B. Gaiffe, J.M. Pierrel, et L. Romary. Informations lexicales dans un dialogue homme-machine multimodal : application au traitement de la coréférence. Dans *Lexique et inférence(s) VII ème Colloque International de Linguistique*, pages 203-226. J.E Tyvaert, Metz, 1991.
- [Gaiffe et al., 1992] B. Gaiffe, L. Romary, et J.M. Pierrel. De la référence à la modélisation de la tâche: vers une certaine idée du dialogue homme-machine, 1992. *Séminaire Dialogue des pôles parole et langage naturel du GDR-PRC Communication Homme-Machine*, Dourdan, 15-16 Avril 1992.
- [Gaiffe et al., 1994] B. Gaiffe, A. Reboul, et L. Romary. Les sn définis : anaphore, anaphore associative et cohérence. Dans *Colloque Anaphores et (in)cohérence*, Université d'Anvers, 1994.
- [Gaiffe et Romary, 1990] B. Gaiffe et L. Romary. Managing Multimodal Information and References in the Multiworks Project. Rapport Multiworks ESP 2105 et rapport CRIN 90R168, 1990.
- [Gaiffe et Romary, 1993] B. Gaiffe et L. Romary. Ce matin, il pleuvait, ou la référence à l'épreuve du temps. *Langages*, (112):203-226, 1993.
- [Gaiffe et Romary, 1995] B. Gaiffe et L. Romary. The treatment of evolving reference in man-machine dialogues. Dans *Evolving reference and anaphora: time and objects*, rédacteur Anne Reboul, Pragmatics and beyond. Benjamins, 1995.
- [Gaiffe, 1992] B. Gaiffe. *Référence et Dialogue homme-machine : vers un modèle adapté au multi-modal*. 1992. Thèse d'Université Henri Poincaré - Nancy I.
- [Gaildrat et al., 1993] V. Gaildrat, N. Vigouroux, R. Caubet, et G. Pérennou. Conception d'une interface multimodale pour un modeleur déclaratif de scènes tridimensionnelles pour la synthèse d'images. Dans *L'Interface des mondes réels et virtuels*, pages 415-424, Montpellier, 22-26 Mars 1993. 2 ème Conférence Internationale.
- [Gardan, 1987] Y. Gardan. *Éléments de C.A.O*, volume 1. Hermès, 1987.
- [Garnham, 1989] A. Garnham. A unified theory of the meaning of some spatial relational terms. *Cognition* 31, pages 45-60, 1989.

- [Garrod et Anderson, 1987] S. Garrod et A. Anderson. Saying what you mean in dialogue: a study in conceptual and semantic co-ordination. Dans *Cognition*, volume 53, pages 181–218. 1987.
- [Garrod et Doherty, 1994] S. Garrod et G. Doherty. Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions. Dans *Cognition*, volume 53, pages 181–215. 1994.
- [Grice, 1979] H. P. Grice. Logic et conversation. *Communications*, (30):57–72, 1979. Traduction française effectuée par O. Ducrot.
- [Hampden-Turner, 1990] C. Hampden-Turner. *Atlas de notre cerveau. Les grandes voies du psychisme et de la cognition*. Les éditions d'Organisation, 1990.
- [Hathout, 1994] N. Hathout. De la sémantique des verbes de déplacement et des prépositions spatiales vers leur syntaxe: cadres formels et théoriques. Manuscrit, 1994.
- [Haton et al., 1991] J.P. Haton, J.M. Pierrel, J. Caelen, et J.L. Gauvain. *Reconnaissance automatique de la parole*. Dunod, 1991.
- [Haton et Haton, 1990] J.P. Haton et M.C. Haton. *L'intelligence artificielle*, volume 2444. Que sais-je?, 1990.
- [Hayes, 1986] P. J. Hayes. Steps towards integrating natural language and graphical interaction for knowledge-based systems. *E.C.A.I.*, (1):456–465, 1986.
- [Hernández, 1994] D. Hernández. Qualitative representation of spatial knowledge. Dans *Lecture Notes in Artificial Intelligence 804*. Springer-Verlag, 1994.
- [Herskovits, 1986] A. H. Herskovits. *Language and Spatial Cognition: An Interdisciplinary Study of the Prepositions in English*. Cambridge University Press, 1986.
- [Hottenroth, 1993] P.M. Hottenroth. Prepositions and object concepts: A contribution to cognitive semantics. Dans *The semantics of prepositions*, pages 179–219. Mouton de Gruyter, 1993. Cornelia Zelinsky-Wibbelt (Editor).
- [Kamp et Reyle, 1993] H. Kamp et U. Reyle. *From discourse to logic*. Kluwer Academic, 1993.
- [Kamp, 1988] H. Kamp. Discourse representation theory: what it is and where it ought to go. Dans *Lecture Notes in Computer Science*, numéro 320. Springer-Verlag, 1988.
- [Kayser, 1988] D. Kayser. Le raisonnement à profondeur variable. *Actes des journées du PRC IA Toulouse, Teknea*, pages 109–136, 1988.
- [Kayser, 1992] D. Kayser. Profondeur variable et sciences cognitives. Dans *Introduction aux sciences cognitives*, pages 195–218. Folio, 1992.
- [Kleiber, 1981] G. Kleiber. *Problèmes de référence: descriptions définies et noms propres*. Klincksiek, 1981.

- [Kleiber, 1990a] G. Kleiber. *La sémantique du prototype, catégories et sens lexical*. P.U.F, 1990.
- [Kleiber, 1990b] G. Kleiber. Quand “il” n’a pas d’antécédent. *Langages*, (97):24–50, 1990.
- [Kleiber, 1991] G. Kleiber. Prototype et prototypes : encore une affaire de famille. Dans *Sémantique et cognition*, pages 103–129. Editions du CNRS, 1991.
- [Klein, 1993] W. Klein. L’expression de la spatialité dans le langage humain. Dans *Images et Langages. Multimodalité et Modélisation Cognitive*, pages 73–85. Colloque interdisciplinaire du CNRS, Paris, 1er et 2 Avril 1993.
- [Koffka, 1935] K. Koffka. *Principles of Gestalt psychology*. New York : Harcourt Brace, 1935.
- [Kohler, 1947] W. Kohler. *Gestalt psychology: an introduction to new concepts in modern psychology*. New York : Liveright Publishing Corporation, 1947.
- [Langacker, 1991] R. W. Langacker. *Concept, Image and Symbol; The Cognitive Basis of Grammar*. Mouton de Gruyter, 1991.
- [Laur, 1993] D. Laur. La relation entre le verbe et la préposition dans la sémantique du déplacement. Dans *La couleur des prépositions*, numéro 110. Larousse, Juin 1993.
- [Leyens, 1992] J. P. Leyens. *Sommes-nous tous des psychologues?*, volume 119. Psychologie et Sciences Humaines, 1992. 292 pages.
- [Mallet, 1992] J. L. Mallet. Discrete smooth interpolation in geometric modelling. *Computer Aided Design Journal*, 24(4), 1992.
- [Marc, 1993] E. Marc. Psychothérapies et développement personnel. *Sciences humaines*, (26):14–20, 1993.
- [Mignot *et al.*, 1993] C. Mignot, C. Valot, et N. Carbonell. An experimental study of future natural multimodal human-computer interaction. *INTERCHI, Amsterdam*, pages 67–68, 1993.
- [Mignot, 1995] C. Mignot. *Usage de la parole et du geste dans les interfaces multimodales, étude expérimentale et modélisation*. 1995. Thèse d’Université Henri Poincaré - Nancy I.
- [Miller et Johnson-Laird, 1976] G. Miller et P. Johnson-Laird. *Language and Perception*. Belknap Press of Harvard University Press, (Cambridge, Massachussets), 1976.
- [Miller, 1956] G. A. Miller. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, (63):81–97, 1956.
- [Milner, 1982] J. C. Milner. *Ordres et raisons de la langue*. Editions du Seuil, Paris, 1982.
- [Moeschler et Reboul, 1994] J. Moeschler et A. Reboul. *Dictionnaire encyclopédique de pragmatique*. Editions du Seuil, Paris, 1994.

- [Montague, 1970] R. Montague. English as a formal language. Dans *Formal philosophy*, rédacteur Thomason. Yale university press, 1970.
- [Morin, 1994] P. Morin. *PARTNER, un système de dialogue homme-machine multimodal pour applications finalisées à forte composante orale*. 1994. Thèse d'Université Henri Poincaré - Nancy I.
- [Mounin, 1972] G. Mounin. *Clefs pour la sémantique*. Seghers, 1972.
- [Mukerjee et Joe, 1990] A. Mukerjee et G. Joe. A qualitative model for space. *AAAI 90 MIT Press Cambridge MA*, pages 721–727, 1990.
- [Mukerjee, 1989] A. Mukerjee. A representation for modelling functional knowledge in geometric structures. Dans *Lecture Notes in Artificial Intelligence 444*, rédacteurs S. Ramani, R. Chandrasekar, et K.S.R Anjaneyulu, pages 192–202. Springer-Verlag, 1989.
- [Nicolle *et al.*, 1997] A. Nicolle, J.M. Pierrel, L. Romary, G. Sabah, A. Vilant, et J. Vivier. *Machine, langage et dialogue*. 1997. À paraître.
- [Nirenburg et Raskin, 1987] S. Nirenburg et V. Raskin. Dealing with space in natural language processing, 1987. In proceedings of the Workshop on Spatial Reasoning and Multi-Sensor Fusion, eds A. Kak and S-S. Chen, 361-370, Los Altos, Calif.: Morgan Kaufmann.
- [Osborn, 1988] A.F. Osborn. *Créativité, l'imagination constructive*. Dunod, 1988.
- [Peroche *et al.*, 1990] B. Peroche, J. Argence, D. Ghazanfarpour, et D. Michelucci. *La synthèse d'images*. Hermès, 1990.
- [Pierart, 1979] B. Pierart. Genèse et structuration des marqueurs de relations spatiales entre trois et dix ans. Dans *Cahiers de l'Institut de Linguistique de Louvain (CILL)*, 5. 1-2, pages 41–59. 1979.
- [Piéron, 1961] H. Piéron. *La sensation*. Que sais-je?, 1961.
- [Pierrel et Romary, 1997] J.M. Pierrel et L. Romary. Quelles références dans les dialogues homme-machine. Dans *Machine, langage et dialogue*, 1997. À paraître.
- [Pierrel, 1987] J.M. Pierrel. *Dialogue Oral Homme-Machine, connaissances linguistiques, stratégies et architectures des systèmes*. Hermès, Paris, 1987.
- [Pierrel, 1992] J. M. Pierrel. Mémoire et contexte dans le traitement automatique du dialogue oral homme-machine, 1992. Journée scientifique, Réseau Cognisciences-Est, Strasbourg, 3 décembre 1992.
- [Pigani, 1995] E. Pigani. Le cerveau, une usine à idées. *Psychologies*, (133):27–30, 1995.

- [Poirier et Lefebvre, 1993] F. Poirier et P. Lefebvre. Une interface multimodale de dialogue homme-machine sous Unix. Dans *L'Interface des mondes réels et virtuels*, pages 161–170, Montpellier, 22-26 Mars 1993. 2^{ème} Conférence Internationale.
- [Pouteau, 1995] X. Pouteau. *Dialogue de commande multimodal en milieu opérationnel : une communication naturelle pour l'opérateur ?* 1995. Thèse d'Université Henri Poincaré - Nancy I.
- [Preparata et Shamos, 1985] F.P. Preparata et M. I. Shamos. *Computational geometry : an introduction*. Springer-Verlag, 1985.
- [Pribbenow, 1993] S. Pribbenow. Computing the meaning of localization expressions involving prepositions: the role of concepts and spatial context. Dans *The semantics of prepositions*, pages 441–470. Mouton de Gruyter, 1993. Cornelia Zelinsky-Wibbelt (Editor).
- [Rastier, 1987] F. Rastier. *Sémantique interprétative*. Formes sémiotiques P.U.F, 1987.
- [Rastier, 1991] F. Rastier. *Sémantique et recherches cognitives*. Formes sémiotiques P.U.F, 1991.
- [Retz-Schmidt, 1988] G. Retz-Schmidt. Various views on spatial prepositions. *A.I Magazine*, 9(2):95–105, 1988.
- [Rock et Palmer, 1991] I. Rock et S. Palmer. L'héritage du gestaltisme. *Pour la Science*, (160):64–70, 1991.
- [Romary et Gaiffe, 1995] L. Romary et B. Gaiffe. Langue, dialogue finalisé et cognition spatiale. Dans *Langage et cognition spatiale*. Editions Masson, 1995. M. Denis (Editeur), à paraître.
- [Romary, 1989] L. Romary. *Vers la définition d'un modèle cognitif pour la représentation du temps dans un système de dialogue homme-machine*. 1989. Thèse d'Université Henri Poincaré - Nancy I.
- [Romary, 1991] L. Romary. Integration of spatial and temporal information produced by a natural language discourse. Dans *Conference Knowledge Modelling and Expertise Transfert*, Sophia Antipolis, April 1991.
- [Romary, 1993] L. Romary. "Mets ça ici" où quand "ici" dépend de "ça". l'interprétation de "ici" dans les énoncés de positionnement. Dans *Workshop : Le Dialogue Homme-Robot en Langage Naturel : Problèmes Psychologiques*, Caen - France, 1993.
- [Rosch, 1978] E. Rosch. Principles of categorization. Dans *Cognition and categorization*, rédacteurs E. Rosch et B. Loyd, pages 27–47. Hillsdale, (N.J.) : L. Erlbaum, 1978.
- [Rosenthal et Frode, 1963] R. A. Rosenthal et K.L. Frode. The effect of experimenter bias on the performance of the albino rat. *Behavioral Science*, (8):183–189, 1963.

- [Sabah, 1989] G. Sabah. *Processus de compréhension*. Hermès, Paris, 1989. Vol 2.
- [Saussure, 1968] F. Saussure. *Cours de linguistique générale*. Payot, 1968.
- [Schang et Romary, 1994a] D. Schang et L. Romary. Frames, a unified model for the representation of reference and space in a man-machine dialogue. *3rd International Conference on Spoken Language Processing (ICSLP-94)*, 1994. Yokohama (Japon).
- [Schang et Romary, 1994b] D. Schang et L. Romary. Framing the world, towards a localised spatial reasoning. *3rd International Conference on the Cognitive Science of Natural Language Processing (CSNLP-94)*, 1994. Dublin (Irlande).
- [Schang, 1991] D. Schang. *Références à des objets multimodaux en langage naturel*. 1991. DEA de l'Université Henri Poincaré - Nancy I.
- [Schang, 1994] D. Schang. Modélisation de la référence et du positionnement spatial dans un système de dialogue homme-machine. *II èmes Rencontres Nationales des Jeunes Chercheurs en Intelligence Artificielle (RJCIA-94)*, 1994. Marseille.
- [Schirra et Stopp, 1993] J. Schirra et E. Stopp. Antlima - a listener model with mental images. *IJCAI*, pages 175–180, 1993.
- [Schirra, 1990] J. Schirra. A contribution to reference semantics of spatial prepositions: The visualization problem and its solution in VITRA, 1990. SFB 314, XTRA n. 75, Université de Saarlandes, Saarbrück, Allemagne.
- [Schirra, 1993] J. Schirra. A contribution to reference semantics of spatial prepositions: The visualization problem and its solution in VITRA. Dans *The semantics of prepositions*, pages 471–515. Mouton de Gruyter, 1993. Cornelia Zelinsky-Wibbelt (Editor).
- [Searle, 1972] J. R. Searle. *Les actes de langage*. Paris, Hermann, 1972.
- [Sergent, 1993] J. Sergent. Aspects neurocognitifs du traitement des visages et du langage verbal et musical. Dans *Images et Langages. Multimodalité et Modélisation Cognitive*, pages 11–28. Colloque interdisciplinaire du CNRS, Paris, 1er et 2 Avril 1993.
- [Shieber, 1986] S. M. Shieber. An introduction to unification based approaches to grammar. Dans *CSLI Lecture notes, Center for the Study of Language and Information, Stanford University*. 1986. Distribué par University of Chicago Press.
- [Sidner, 1986] C. Sidner. *Focussing in the comprehension of definite anaphora*, pages 363–394. Readings in natural language processing. Morgan Kauffman Publisher, 1986. Edité par B. Grosz, K. S. Jones, B. L. Webber.
- [Sondheimer, 1976] N. K. Sondheimer. Spatial reference and natural-language machine control. *International Journal of Man-Machine Studies*, (8):329–336. 1976.
- [Souvay, 1992] G. Souvay. *DIAPASON: un environnement de développement pour l'intégration d'une entrée vocale dans des applications de type commande de machine*. 1992. Thèse d'Université Henri Poincaré - Nancy I.

- [Sperber et Wilson, 1989] D. Sperber et D. Wilson. *La pertinence. Communication et Cognition*. Editions de Minuit, 1989.
- [Stopp, 1993] E. Stopp. Konstruktion dreidimensionaler mentaler Bilder aus sprachlichen Szenenbeschreibungen. Memo 60, SFB 314, Universität des Saarlandes - Saarbrücken, 1993.
- [Taylor, 1993] J. R. Taylor. Prepositions: Patterns of polysemization and strategies of disambiguation. Dans *The semantics of prepositions*, pages 151–175. Mouton de Gruyter, 1993. Cornelia Zelinsky-Wibbelt (Editor).
- [Vandeloise, 1986] C. Vandeloise. *L'espace en français*. Editions du Seuil, Paris, 1986.
- [Vandeloise, 1988] C. Vandeloise. Les usages spatiaux statiques de la préposition "à". Dans *Cahiers de lexicologie*, volume 53, pages 119–148. 1988.
- [Vieu, 1991] L. Vieu. *Sémantique des relations spatiales et inférences spatio-temporelles : une contribution à l'étude des structures formelles de l'espace en langage naturel*. 1991. Thèse d'Université Paul Sabatier Toulouse.
- [Vilnat, 1997] A. Vilnat. Quels processus pour les dialogues homme-machine. Dans *Machine, langage et dialogue*, 1997. À paraître.
- [Vivier, 1997] J. Vivier. Pour une psychologie du dialogue homme-machine. Dans *Machine, langage et dialogue*, 1997. À paraître.
- [Wahlster, 1987] W. Wahlster. One word says more than a thousand pictures, on the automatic verbalization of the results of image sequence analysis system. *T.A informations, Bulletin semestriel de l'ATALA*, (2):57–70, 1987.
- [Wazinski, 1991] P. Wazinski. Generating spatial descriptions for cross-modal references. Rapport Technique TM-91-11, DFKI, Saarbrücken, 1991.
- [Wazinski, 1993] P. Wazinski. Graduated topological relations. Memo 54, SFB 314, Universität des Saarlandes - Saarbrücken, 1993.
- [Wertheimer, 1912] M. Wertheimer. Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie*, (61):161–265, 1912.
- [Wittgenstein, 1953] L. Wittgenstein. *Philosophical investigations*. Mac Millan, New-York, 1953.
- [Woods, 1970] Woods. Transition network grammars for natural language analysis. *CACM*, pages 591–606, 1970.
- [Woollams et al., 1990] S. Woollams, M. Brown, et K. Hume. *Analyse transitionnelle*. Imprimerie Nationale Jean Peutey - Genève, 1990.

Glossaire

Dans ce glossaire, le lecteur trouvera une définition de termes techniques, issus notamment de la linguistique, de la psychologie et de l'informatique, qui n'ont pas toujours été définis dans ce mémoire. Les mots en italiques sont eux-mêmes des entrées du glossaire.

Anaphore : il y a anaphore chaque fois que l'interprétation référentielle d'une expression n'est possible que par le truchement d'une autre présente dans le *contexte*.

Application : au sens informatique du terme : ensemble de compétences logicielles permettant de réaliser des *tâches*.

Axiologie : point de vue référentiel adopté par un discours pour exprimer une opposition élémentaire [Gaiffe, 1992]. Il s'agit d'une opposition séparant les éléments d'un ensemble aux deux extrémités d'un axe d'opposition étiqueté par une propriété *P* satisfaite d'un côté et fautive de l'autre.

Cataphore : *référence* en avant (exemple : "*Quand il ouvrit la porte, Pierre vit...*").

Cible : entité localisée par rapport au *site* dans le formalisme de Vandeloise.

Cognition : la cognition, ou plus généralement la connaissance, est l'objet d'une science récente, la psychologie cognitive, qui ne s'intéresse plus, ou plus uniquement, aux pathologies mentales, mais à la façon dont fonctionne l'esprit humain.

Compositionnalité : le *sens* d'un message se déduit du *sens* de ses composants.

Contexte : environnement *linguistique* d'un terme ou d'un *énoncé*.

Co-référence : il y a relation de co-référence entre deux unités référentielles A et B quand elles se trouvent avoir la même *référence*, ce qui peut arriver sans que l'interprétation de l'une soit affectée par l'interprétation de l'autre.

Déictique : marqueur *linguistique* faisant *référence* à un élément extra-linguistique appartenant à la situation d'énonciation (exemple : "*ici*", "*là*", "*maintenant* ").

Dialogue : échange de symboles, effectué au moyen d'un médium approprié, entre

deux ou plusieurs entités. La signification des symboles est fondée sur l'hypothèse d'un référentiel d'interprétation commun partagé par les participants. en pratique, chaque participant possède un référentiel propre lié entre autres à sa culture, son éducation et son état émotionnel (d'après [Morin, 1994]).

Dialogue multi-modal: *dialogue* dans lequel une intervention s'exprime, à un moment donné, au moyen de plusieurs modes de communication (parole, geste, etc).

Discours: *énoncé* ou suite d'*énoncés*, oraux ou écrits, formant une unité.

Ellipse: Omission syntaxique ou stylistique de un ou plusieurs mots que l'esprit humain supplée de façon plus ou moins spontanée.

Énoncé: message composé dans une *langue*.

Fonction: action, rôle caractéristique d'un élément, d'un organe dans un ensemble.

Composante fonctionnelle: relative aux fonctions, ensemble des connaissances extralinguistiques que partagent les locuteurs d'une même *langue*.

Heuristique: règle que l'on a en général intérêt à utiliser. Elle aide souvent à trouver la solution, bien qu'on n'ait aucune certitude sur sa validité dans tous les cas possibles.

Historique du dialogue: structure permettant de garder en mémoire la forme langagière des différentes interventions des interlocuteurs et l'organisation logique de l'ensemble du dialogue.

Hyperonyme: terme générique (opposé à hyponyme), "*insecte*" est l'hyperonyme de "*papillon*", "*mouche*", etc.

Implicature: le terme d'implicature désigne certaines conclusions que l'on peut tirer d'énoncés sans que la relation entre ces conclusions et les énoncés en question puisse se ramener à la relation logique d'implication. On distingue classiquement les *implicatures conventionnelles* des *implicatures conversationnelles*. Les premières sont obtenues à partir des termes linguistiques eux-mêmes et sont très proches des présuppositions ou des présomptions lexicales. L'énoncé "*Jean a cessé de fumer*" produit l'implicature conventionnelle que *Jean fumait*". Les implicatures conversationnelles sont obtenues par inférence non démonstrative à partir des énoncés et des maximes conversationnelles. Ainsi, si A propose du café à B et que B lui répond que le café l'empêche de dormir, l'énoncé de B a pour implicature conversationnelle que B ne veut pas de café (d'après [Moeschler et Reboul, 1994]).

Langage: 1) activité consistant à communiquer grâce à une *langue*, 2) ensemble des *langues*.

Langage naturel : représente l'ensemble des *langues* du monde.

Langage artificiel : langage reposant sur des axiomes, des lois, des règles de formation des énoncés.

Langue : système de *signes* commun à un groupe social.

Linguistique : science qui a pour objet l'étude de la *langue*.

Objectivisme : attitude pratique qui consiste à s'en tenir aux données contrôlables par les sens, à écarter les données subjectives.

Orientation générale : *ressemblance de famille* qui compte pour principaux traits la direction frontale, la direction du mouvement, la ligne du regard, la direction dans laquelle sont dirigés les autres organes : la perception, l'odorat, l'ouïe....

Orientation latérale : il s'agit d'une *ressemblance de famille* dont les principaux traits sont la direction latérale et la perpendiculaire à l'*orientation générale*.

P.é.c : principe d'économie cognitive.

Pragmatique : étude des *signes* en situation.

Effet Pygmalion : influence des attentes du sujet envers son partenaire qui ont un effet déterminant sur les résultats expérimentaux effectifs.

Référence : la référence est la relation qui unit une expression de la *langue* (dite en général expression référentielle) en emploi dans un énoncé et l'objet dans le monde que cette expression désigne.

Référent : être, objet, événement ou propriété de la réalité (vraie ou imaginaire) qu'un *signe* évoque (exemple : la plante que le mot "rose" évoque).

Ressemblance de famille : concept représenté par la combinaison des traits qui le caractérisent, sachant que chaque trait n'est pas toujours nécessaire ni suffisant.

Sciences cognitives : elles ont pour objet de décrire, d'expliquer et le cas échéant de simuler les principales dispositions et capacités de l'esprit humain : langage, raisonnement, perception coordination motrice, planification, etc. Les principales disciplines qui en font partie sont la linguistique, la psychologie, l'anthropologie, les neurosciences, l'intelligence artificielle, la logique et la philosophie (d'après [Andler, 1992]).

Sémantique : science dont l'objet est l'étude du *sens*.

Sens : ce que veut dire un mot, ou un *énoncé*, prononcé ou écrit.

Signe : union d'un *signifiant* et d'un *signifié*.

Signifiant : forme extérieure, face perceptible d'un *signe*; s'oppose à *signifié*.

Signifié : *sens* d'un *signe*.

Site : entité localisatrice dans le formalisme de Vandeloise.

Syntagme : dans un *énoncé*, groupe de mots constituant une unité syntaxique (exemple : "le chien de Blandine").

Syntaxe : ensemble des règles de la *langue* qui permettent de construire des phrases à partir des mots et de leurs catégories.

Tâche : objet du face-à-face homme-machine. La tâche consiste pour l'opérateur humain en l'atteinte d'un but, ce but représentant la fin d'une étape particulière de son activité.

Nom : **SCHANG**

Prénom : **Daniel**

DOCTORAT de l'UNIVERSITE HENRI POINCARÉ, NANCY-I
en **INFORMATIQUE**

VU, APPROUVÉ ET PERMIS D'IMPRIMER

Nancy, le 20 JAN 1997 UHP 010/97

Le Président de l'Université



Résumé

Cette thèse présente un ensemble de recherches dont l'objectif est de contribuer à la mise en œuvre de systèmes de dialogue homme-machine possédant un comportement aussi naturel que possible. Ces travaux portent sur le domaine de la référence aux objets pour des dialogues homme-machine multimodaux. Plus spécifiquement, dans le cadre d'une coopération entre expressions langagières et visualisation graphique d'un espace de désignation, nous nous intéressons aux références spatiales s'exprimant à l'aide de prépositions spatiales. La réflexion proposée s'articule autour des quatre axes principaux suivants :

- dans un premier temps, nous fournissons la trame de cette thèse qui se fonde sur le triptyque suivant : dialogue, contexte et principe d'économie cognitive,

- nous cherchons dans un second temps, à élaborer un modèle qui puisse effectuer des références spatiales portant sur un environnement accessible visuellement à un utilisateur,

- le caractère par trop *ad hoc* du modèle précédent nous incite, dans le cadre des sciences cognitives, à élaborer un modèle de représentation de l'espace qui permet de considérer sous un même angle des phénomènes liés au langage, à la représentation des connaissances et au raisonnement spatial et ce, à l'aide de la notion générale de cadre. Cette notion est instanciée pour le problème spécifique des énoncés de référence spatiaux sous la dénomination de cadre de référence dont nous décrivons les tenants et les aboutissants,

- l'aboutissement de ce travail permet de donner certaines pistes en vue de l'intégration du modèle proposé dans un système opérationnel.

Mots-clés: dialogue homme-machine, langage naturel, référence, modélisation spatiale.

