



Langue, dialogue finalisé et cognition spatiale

Laurent Romary, Bertrand Gaiffe

► **To cite this version:**

Laurent Romary, Bertrand Gaiffe. Langue, dialogue finalisé et cognition spatiale. Michel Denis. Langage et Cognition Spatiale, Masson, 19p., 1997. inria-00488852

HAL Id: inria-00488852

<https://hal.inria.fr/inria-00488852>

Submitted on 3 Jun 2010

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.

LANGUE, DIALOGUE FINALISÉ ET COGNITION SPATIALE

Laurent Romary et Bertrand Gaiffe
CRIN-CNRS et INRIA Lorraine
Bâtiment Loria, B.P. 239, 54506 Vandœuvre lès Nancy
tél. 83 59 20 37
fax 83 41 30 79
e-mail : {romary,gaiffe}@loria.fr

1. Dialogue et espace

L'objectif de ce chapitre est d'aborder l'espace sous l'angle de la communication homme-machine pour mettre en évidence ce qui peut caractériser les mécanismes de compréhension d'expression de l'espace, ainsi que les opérations sous-jacentes de perception et de raisonnement. Nous allons essayer de montrer comment la perspective de réaliser des systèmes opérationnels de dialogue homme-machine permet d'intégrer la modélisation de l'espace dans une perspective plus large intégrant de nombreux paramètres liés à tout acte de communication.

Quel espace pour le dialogue homme-machine ?

Parmi les applications possibles des systèmes de dialogue homme-machine utilisant une entrée langagière orale ou écrite, on peut distinguer deux grandes classes : d'une part, celle des systèmes « distants » auxquels l'utilisateur accède via le téléphone par exemple, et d'autre part celle où l'utilisateur est en contact direct avec le système, ce qui permet en particulier de lui fournir une représentation graphique des objets qu'il est susceptible de manipuler au cours de la réalisation de sa tâche. Dans ce chapitre, nous ne nous intéresserons qu'à cette seconde classe pour laquelle l'analyse d'expressions spatiales semble effectivement la plus pertinente¹.

Dans cette perspective, l'écran d'un ordinateur donne accès à l'espace pertinent pour l'application. Pour autant, la surface rectangulaire constituée par cet écran n'est pas nécessairement le seul espace de l'application. Ainsi, dans la mesure où, en pratique, l'écran sert de support à la représentation d'un monde particulier lié à l'application, l'espace de cette application intervient également. Ainsi, dans le cadre d'une expérimentation récente² où des utilisateurs devaient (à travers une simulation de dialogue homme-machine de type « Magicien d'Oz ») aménager des appartements, la situation a fait apparaître deux types d'espace :

- a. celui de l'écran dans lequel on peut désigner un fauteuil comme étant "celui de droite",
- b. celui de l'appartement dans lequel un fauteuil peut être disposé près de la cheminée.

Dans ce second espace, la fonctionnalité des meubles interagit avec les expressions spatiales - ainsi que Vandeloise l'a observé par ailleurs (Vandeloise 1986) - et justifie des expressions telles que *devant la cheminée* ou *devant le fauteuil*.

¹ Dans le cas de conversations téléphoniques, la dimension spatiale est généralement réduite aux seules références déictiques centrées sur les interlocuteurs (e.g. *Ici, il fait un temps magnifique*).

² Il s'agit d'une étude menée conjointement par le Centre de Recherche en Informatique de Nancy et le CERMA (Dauchy, et al. 1993).

Dans le cadre de certaines applications de commande, il est également possible d'avoir affaire à des espaces plus abstraits en ce qu'ils ne correspondent pas à une projection (ou une métaphore) de notre espace de perception habituel. Ainsi, dans un système de dialogue de commande dédié à des opérateurs sonar³ par exemple, l'espace représenté sur l'écran est en fait un espace de (temps x angles). Par ailleurs, pour des raisons physiques (liées au sonar), la présence d'un objet se reflète en général par deux traces symétriques : on ne sait pas si l'objet repéré est à gauche ou à droite. Bien évidemment, dans un tel contexte, des expressions telles que *la piste de droite* sont susceptibles d'un traitement particulier selon le contexte.

Enfin, plusieurs organisations spatiales peuvent cohabiter sur un même écran, le système de dialogue dédié au contrôle aérien étudié (et réalisé) par Gilles Souvay dans le cadre du contrat Esprit Roars (Souvay et Pierrel 1995) mêle ainsi, un espace dans lequel apparaissent les « strips » des vols (représentant une métaphore des rubans de papier actuellement utilisés par les contrôleurs aériens) et un espace de pistes radar. On peut alors envisager des expressions telles que :

le strip du haut voire *le vol du haut* dans l'espace de représentation des strips,
la piste du haut voire *le vol du haut* dans l'espace des pistes radar représentées,
et enfin *le plus haut* dans l'espace « réel » du ciel où évoluent des avions, appréhendable par l'altitude affichée sur les pistes radar.

Ce dernier type d'expression est d'autant plus plausible que les pistes radars peuvent sembler avoir des intersections dans leur projection en deux dimensions sans en avoir dans l'espace en trois dimensions et que l'objectif d'un contrôleur aérien est bien entendu d'éviter des intersections (et donc des collisions...) dans l'espace à trois dimensions.

Il est donc clair que l'application sur laquelle le système de dialogue est mis en œuvre apporte ses propres contraintes quand à la représentation de l'espace à considérer. Pour autant, des constantes semblent se dégager dans les possibilités de référer de façon spatiale aux objets de l'application. Tout au long de ce chapitre, nous utiliserons ainsi des exemples qui apparaîtront probablement éloignés des contingences d'une application effective[@] mais, nous semble-t-il, sont représentatifs des difficultés que l'on peut rencontrer dans des contextes réels tels que ceux brièvement mentionnés ci-dessus.

Deux logiques d'interface possibles : outils ou partenaire

Le concepteur de l'interface d'un système de dialogue de commande destiné à être utilisé dans les conditions que nous avons mentionnées doit faire un choix entre deux logiques d'utilisation de son système (Romary, et al. 1993, Pouteau, et al. 1994). Soit il entre dans une logique d'outil, soit dans une logique de communication.

Le premier de ces choix consiste à donner, pour chacune des actions possibles de la part de l'utilisateur et pour chaque objet susceptible d'être argument d'une commande, un moyen de désignation non ambigu et entièrement codé. Typiquement, ce mode d'interaction donne lieu, en ce qui concerne la gestion des gestes à la notion d'objet réactif.

En terme spatiaux, ce choix conduit à une granularité de l'espace (pour la désignation) qui se trouve être fixée *a priori*. Ainsi, sur un Macintosh, la case de fermeture d'une fenêtre est globalement homogène pour la désignation à l'aide de la souris. Ce type d'approche, lorsqu'elle

³ Le développement d'un tel système a fait l'objet d'un contrat de collaboration entre le CRIN et l'entreprise Thomson DASM, sous l'égide de la DRET.

utilise des mots de la langue, par exemple au moyen d'un système de reconnaissance de parole aboutit à la notion de langage artificiel éventuellement à consonance naturelle. En effet, l'interprétation du langage doit alors être telle que chaque commande possible soit accessible par le biais d'un ensemble prédéterminé de formes. Dans tous les cas, l'utilisateur agit directement sur l'interface et se place donc dans une logique du « faire ».

La seconde approche vise des systèmes « partenaires ». On entre alors dans une logique du "faire-faire" (par le système), et l'utilisateur doit communiquer des buts, à charge pour le système de trouver des moyens de les réaliser. C'est dans cette seconde perspective que nous nous placerons tout au long de ce chapitre. Dans la mesure où cette fois, le système doit interpréter les énoncés de l'utilisateur en termes d'intention, il pourra y avoir dialogue avec le système et structuration inter-énoncé. De façon plus générale, l'interprétation sera contextuelle et dépendra en particulier d'une structuration de l'espace si possible partagée par les interlocuteurs.

2. Différents modes de référence à l'espace

Dans le type de situation de communication auquel nous nous intéressons, puisque l'utilisateur est physiquement présent devant une représentation des objets, il semble naturel de l'autoriser à désigner les objets qu'il souhaite modifier à la fois de façon langagière et gestuelle. Depuis plusieurs années déjà, des dispositifs matériels tels que des gants de désignation, des écrans tactiles ou des tablettes graphiques permettent d'approximer l'utilisation de gestes de désignation tels qu'on peut les pratiquer dans des conversations entre humains. Notre propos ne vise pas à nous demander si de tels dispositifs permettent des gestes véritablement naturels ou pas, ni dans des considérations techniques tenant à la mise en oeuvre informatique de ces dispositifs. Nous supposerons disposer d'une approximation raisonnable des gestes de désignation. Dans une telle perspective, isoler une classe particulière d'expressions de la référence à l'espace sous prétexte que les éléments de cette classe contiennent des "mots spatiaux" telles que des prépositions de lieu ou des déictiques tels que "ici" semble peu pertinent. Directement ou indirectement, la plupart des expressions référentielles employées viseront un élément représenté dans l'espace visuel partagé par l'utilisateur du système de dialogue et ce système lui-même. Par conséquent, les expressions spatiales, au sens d'expressions contenant des "mots spatiaux", seront en concurrence avec des expressions habituellement considérées comme non spatiales. Ainsi, dans une situation telle que celle présentée figure 2.1 ci-dessous :



Figure 2.1

On peut faire référence au triangle noir par l'une quelconque des expressions suivantes :

- *le triangle noir*
- *ce triangle* (en le montrant)
- *le triangle de gauche.*

Il nous semble que ce soit l'ensemble de ces expressions qu'il faille globalement considérer si l'on souhaite comprendre la spécificité du dernier énoncé au regard des deux autres. De plus, dans la situation de la figure 2.1, il nous semble que même la première expression peut être vue comme ayant une incidence d'ordre spatial. L'interlocuteur étant amené à juger la couleur des triangles a toutes chances de considérer globalement l'ensemble des deux triangles pour autant qu'ils soient suffisamment proches. Nous verrons d'ailleurs dans la suite de ce chapitre qu'inversement, certaines expressions spatiales peuvent se trouver justifiées par des caractéristiques non directement spatiales des objets désignés. C'est donc l'ensemble des caractéristiques d'ordre visuel qui doit être considéré et pas la seule position des objets. Enfin, le geste de désignation opérant lui aussi dans cet espace visuel, il doit être pris en considération sinon comme expression spatiale, du moins pour les conséquences qu'il peut avoir quand à l'emploi d'expressions spatiales dans les énoncés suivants.

Rôle privilégié de la langue naturelle

Pour désigner un objet dans un environnement visuel, la langue « naturelle » offre une variété de moyens sans comparaison (au moins au premier abord) avec ce que permet le geste. Outre les déterminants eux-mêmes (le, les, un, des...), la langue utilise des catégories nominales qui permettent de un niveau de granularité correspondant à la nature de l'objet visé. Ainsi, là où un geste pourrait être ambigu et désigner aussi bien un triangle complet que plus particulièrement l'un de ses segments, l'expression *le triangle...* lèvera l'ambiguïté. Par ailleurs, un système de dialogue étant dédié à une application particulière, la notion de type s'impose naturellement comme la classe des objets qui supportent un même ensemble d'opérations, or les catégories langagières s'adaptent en général à cette notion de type : il n'existe pas forcément de mot pour l'ensemble des objets déplaçables, mais on a en général des mots permettant de désigner les sous-catégories composant cet ensemble. D'ailleurs, pour autant que l'application ait des utilisateurs professionnels, il existe le plus souvent un langage spécialisé (ou opératif selon la terminologie de Falzon (Falzon 1986)) dans lequel ces catégories existent.

Quelle place pour le geste ?

Contrairement à la parole, faculté humaine tout entière dédiée à la communication, le geste est associé à de nombreuses fonctions permettant d'agir sur un environnement, mais aussi de recueillir des données par l'intermédiaire du sens du toucher en particulier. En reprenant la classification de (Cadoz 1992), nous pouvons distinguer d'une part les fonctions ergative et épistémique (perceptive) du geste qui impliquent un contact direct avec l'environnement, et d'autre part la fonction sémiotique qui elle, correspond à une véritable communication en direction d'un interlocuteur humain ou non.

Comme nous l'avons mentionné précédemment, seuls les gestes sémiotiques nous intéresseront ici. De façon plus précise, nous nous limiterons à des gestes de désignation d'objets. Ce type de geste peut difficilement apparaître indépendamment de tout énoncé langagier, et ce pour deux raisons principales :

- d'une part, le geste seul n'impose pas de granularité spatiale et a donc toute chance d'être ambigu. Ainsi, un geste de désignation dans un texte peut tout aussi bien désigner un caractère qu'un mot une ligne ou un paragraphe.

- d'autre part, dans des dialogues de commande, un énoncé doit aboutir à (ou participer à) un but, c'est-à-dire la spécification d'une modification d'un ou de plusieurs objets. Le geste de désignation ne peut que désigner l'objet, la modification devant alors être exprimée linguistiquement. Dès lors, et pour des raisons purement syntaxiques, il faudra aussi marquer linguistiquement l'argument de ce prédicat et le faire coïncider d'une manière ou d'une autre avec le geste.

Concrètement, dans la mesure où les gestes que nous considérons co-occurrent avec des énoncés langagiers, les conditions d'emploi du geste de désignation dépendent également des conditions d'emploi des expressions langagières qui leur sont liées. Ainsi, il est clair que l'exemple prototypique d'expression faisant coopérer langue et geste sera du type 'ce N' + geste, où N est un nom. Au contraire, une expression du type 'le N' accompagnée d'un geste nécessitera un examen plus approfondi, pour expliquer notamment pourquoi ses conditions d'usages semblent être très restrictives. Notons tout de même que l'adjectif démonstratif n'impose pas nécessairement la cooccurrence avec un geste de désignation, puisqu'il peut aussi être la marque d'une reprise strictement anaphorique comme dans l'exemple suivant :

Colorie la fenêtre de gauche et mets cette fenêtre en bas de l'écran.

Globalement, nous ne considérons donc, d'une part, que des expressions purement langagières et d'autre part des expressions langagières accompagnées de gestes qui seront essentiellement des expressions démonstratives en ce qui concerne la référence à des objets. D'autres expressions peuvent bien sûr intervenir, telles que le déictique *ici* par exemple pour désigner des lieux (Romary, 1993), ou le démonstratif *ça*. De notre point de vue, leur rôle n'est pas fondamentalement différent des autres expressions que nous traitons dans ce chapitre.

3. Nécessité d'une structure préalable

Après avoir dressé un inventaire des possibilités offertes à un locuteur de faire référence aux objets qui lui sont présentés graphiquement dans le cadre d'un système de dialogue auquel on a adjoint une composante de visualisation, nous allons essayer de voir plus précisément quelles sont les conditions d'usage de tel ou tel mode de référence en fonction des indications contextuelles fournies par l'organisation spatiale des objets, mais aussi, comme nous le verrons, par les contraintes inhérentes à la tâche à réaliser. Bien que nous nous placions dans l'objectif de *comprendre* les expressions référentielles adressées au système de dialogue, nous ne pouvons négliger le point de vue du locuteur qui va choisir la meilleure expression possible pour faire identifier par son interlocuteur l'objet sur lequel il souhaite agir, pour en modifier les propriétés par exemple. D'une certaine manière, nous allons donc chercher, malgré le caractère très spécifique de nos réflexions, à ne pas entrer en contradiction avec des principes pragmatiques plus généraux tels que ceux énoncés par la théorie de la pertinence (Sperber et Wilson 1986). En quelques mots, il s'agit pour un locuteur de s'assurer que son énoncé produit un effet pragmatique maximal (résolution des références et des implicatures conversationnelles, effet perlocutoire le plus conforme à l'acte illocutoire) pour une charge cognitive minimale, c'est à dire en limitant le nombre d'inférences nécessaires de la part de l'auditeur. Dans la mise en œuvre de systèmes de dialogue, un tel principe de pertinence peut éventuellement être poussé à l'extrême pour n'autoriser, dans tel ou tel contexte, que l'utilisation de la ou des expressions référentielles les plus pertinentes. Bien qu'un peu limitative, cette attitude permet de compenser en particulier certaines limitations des modules de reconnaissances de parole que nous utilisons en entrée de nos systèmes.

De façon immédiate, on peut considérer que le locuteur donnant à identifier quelque chose doit faire usage des propriétés de l'objet visé pour qu'il puisse être distingué des autres objets de la tâche. Suivant la situation graphique dans laquelle on se trouve, il existe a priori trois moyens principaux d'effectuer une désignation de ce type :

- a) le locuteur peut directement utiliser le nom de la catégorie de l'objet. Dans la majorité des cas, cette mention est d'ailleurs indispensable, mais son usage seul va correspondre aux descriptions définies simples du type *le triangle* ;
- c) il peut montrer directement l'objet, en complément d'une expression démonstrative du type *ce triangle*.
- c) il peut, enfin, utiliser la localisation de l'objet sur l'espace de visualisation, ceci correspondant à des expressions du type *le triangle de gauche* ;

La difficulté qui apparaît est alors de caractériser précisément la nature des contextes graphiques correspondant, pour être réellement amené à comprendre et prédire le comportement des expressions correspondantes. Pour dépasser le cas trivial de l'objet seul sur un écran⁴ nous pouvons observer trois situations où les expressions, *le triangle*, *ce triangle*, *le triangle de gauche* sont respectivement énoncées dans le contexte des figures 3.1, 3.2 et 3.3.

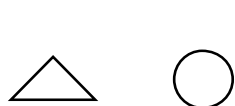


Fig. 3.1 : *le triangle*

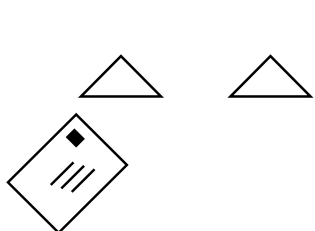


Fig. 3.2 : *ce triangle*

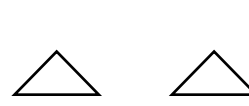


Fig. 3.3 : *le triangle de gauche*

Bien que le cas de la figure 3.3 semble le plus naturel, puisque l'acte de référence s'appuie sur une propriété externe à l'objet dont l'interprétation passe par un point de référence (ici, l'autre triangle), il est relativement facile de montrer que la présence d'un autre objet en 3.1 et 3.2 n'est pas complètement indifférent. Ainsi, une situation telle que celle présentée figure 3.4 n'est pas appropriée, pour des raisons qui sont classiquement analysées comme des ruptures des contraintes d'unicité associées aux descriptions définies. De la même façon, il semblerait maladroit, dans le cas de la figure 3.1, soit de dire *ce triangle* associé à un geste ou *le triangle de gauche*. Pour ce dernier cas, une expression du type *la figure de gauche* serait vraisemblablement une meilleure alternative à *le triangle*⁵.

⁴ il est important de signaler que la situation n'est pas si triviale que cela. Nous nous limitons volontairement ici aux références spatiales portant sur l'univers de la tâche. Vu de façon plus large, même un objet seul à l'écran peut, du point de vue de l'utilisateur, être contrasté avec d'autres éléments de son environnement, que ce soit graphiquement (ex. : *Mets le triangle en vert, il se détachera mieux du fond*) ou temporellement (ex. : *Garde ce triangle*, dans une tâche de présentation successive de figures géométriques). A l'extrême, dans un environnement graphique où il n'y aurait qu'un seul objet, celui-ci jouerait un rôle proche d'un curseur (cf. la tortue Logo) auquel on ne ferait plus explicitement référence. Les énoncés ne porteraient alors que sur l'action (ex. *Vers la gauche. Encore un peu. Stop.*).

⁵ On observera que les deux expressions ne correspondent pas exactement à la même intention communicative. *Le triangle* représente clairement une implication plus fine du locuteur dans la tâche, par rapport à *la figure de gauche* qui caractérise uniformément l'ensemble des objets. De fait, dans une tâche de



Ex. 3.4 : ? le triangle

La difficulté de mettre en œuvre une composante référentielle dans un système de dialogue apparaît ici immédiatement dans la complexité des comportements observés pour les différents types d'expressions que nous avons considérées. Le danger est bien de multiplier les modules de calculs qui vont être activés en fonction de l'expression qui est rencontrée, ce qui crée d'importantes difficultés de gestion comparative des résultats obtenus et donc d'intégration dans le cadre d'une architecture globale de système. Il nous semble donc primordial de proposer un ensemble de mécanismes de fonctionnement de ces expressions faisant référence à l'espace qui repose globalement sur des principes similaires.

A ce stade nous proposons d'analyser l'ensemble des exemples précédents sur la base d'une hypothèse *contrastive*, c'est-à-dire faisant reposer l'interprétation des expressions référentielles sur un mécanisme qui isole un référent, bien plus qu'il ne l'identifie à partir de ses caractéristiques propres. Plus précisément, cette hypothèse peut s'affiner de façon spécifique pour les trois classes d'expressions que nous avons relevées :

1. une description définie en 'le N' réalise un contraste entre ce qui peut être catégorisé en N de ce qui ne le peut pas ;
2. un groupe nominal démonstratif en 'ce N' opère un contraste entre un élément de type N particulier (en position focale) et d'autres éléments de type N ;⁶
3. une description spatiale (ex. *le triangle de gauche*) opère un contraste à l'intérieur d'une catégorie de référence, pour isoler l'élément répondant le plus aux contraintes exprimées par le complément de nom.

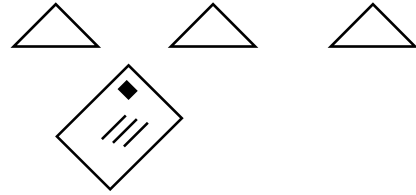
Une observation importante est que la structure contrastive que nous faisons correspondre aux différentes expressions spatiales doit d'une certaine manière être préalable à l'interprétation de celles-ci. Il faut qu'en quelque sorte l'auditeur soit à même de projeter sur l'espace visuel partagé un schéma contrastif indépendamment de l'expression qui à elle seule ne peut pas nécessairement l'imposer. Pour s'en convaincre, il suffit d'observer que dans les cas des figures 3.4.a et 3.5.a, il est particulièrement difficile d'isoler les deux triangles de gauche, alors que la chose est rendue beaucoup plus facile par le contraste visuel existant entre ceux-ci et celui de droite dans le cas des figures 3.4.b et 3.5.b.

dessin telle que celle utilisée par N. Oskan (1994) pour observer des dialogues homme-homme finalisés, ce dernier type de référence est rencontrée de façon beaucoup plus rare.

⁶ Bien que notre analyse soit guidée par des contraintes de mise en œuvre de systèmes de dialogue et ne se veuille donc pas avoir une portée linguistique générale, elle semble faire écho à des travaux tels que ceux de F. Corblin (Corblin 1987), notamment en ce qui concerne le démonstratif. Un approfondissement pluridisciplinaire serait semble-t-il nécessaire pour confronter les perspectives.



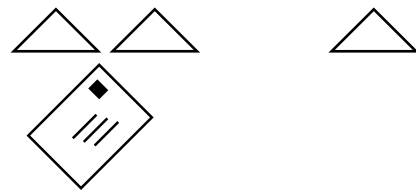
Ex. 3.4 .a : ?les triangles de gauche



Ex. 3.5.a : ?ces triangles



Ex. 3.4 .b : les triangles de gauche



Ex. 3.5.b : ces triangles

Dès lors, il est possible de réinterpréter la contrainte d'unicité généralement associée à l'analyse des descriptions définies comme une contrainte plus large sur l'existence d'une structure où il est possible d'isoler un ou des référents le long d'une propriété catégorisante ou sous-catégorisante compatible avec la structure correspondante. En cela, nous défendons une position duale de celle mise en avant par Robert Dale (Reiter et Dale 1992, Dale 1995) pour la génération d'expressions référentielles, puisque celui-ci insiste sur le fait que ce n'est pas tant la possibilité d'accéder à un référent unique qui permet de juger de la pertinence d'une expression référentielle, que sa capacité à réaliser un véritable effet de distinction dans un ensemble contextuel courant. Pour notre part, il nous apparaît indispensable d'élargir cette notion, en particulier par le biais de l'étude de la référence à l'espace, pour qu'au delà d'une notion d'ensemble de référents, classiquement utilisée pour représenter les cadres contrastifs (Gaiffe, et al. 1992), il soit possible de définir une structure référentielle plus proche de ce que semble être le comportement cognitif humain vis-à-vis des objets perçus et manipulés dans le discours.

Ainsi, dans le cas d'expressions définies spatiales (par exemple *le triangle de gauche*, figure 3.6), on observe en fait deux niveaux successifs de discrimination qui semblent correspondre à deux niveaux de structures imbriquées. Dans un premier temps, la tête nominale effectue un filtrage sur la catégorie permettant de n'isoler que les triangles parmi les objets considérés. Puis, sur la base des triangles ainsi obtenus, le complément de nom projette à son tour la structure nécessaire à son interprétation.



Ex. 3.6 : le triangle de gauche

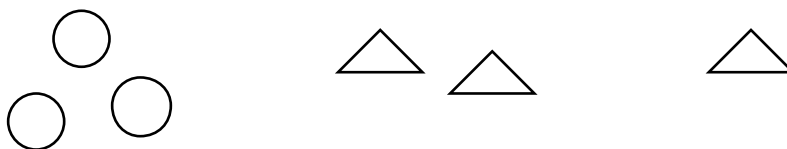
Il est relativement facile de montrer que l'imbrication des structures s'opère bien dans l'ordre que nous avons énoncé et donc que l'information spatiale n'est interprétable que dans un deuxième temps. Ainsi l'exemple 3.7 s'interprète dans les mêmes conditions que l'exemple 3.4.b, alors que l'exemple 3.8 pose autant de problèmes d'interprétation que 3.4.a⁷. De la même façon, il est clair qu'il serait impossible que le calcul aboutisse à des triangles dans l'exemple 3.9 si l'instruction spatiale prenait le pas sur l'instruction de catégorie, ce qui n'est manifestement pas le cas.



Ex. 3.7 : les triangles de gauche



Ex. 3.8 : ? les triangles de gauche



Ex. 3.9 : les triangles de gauche

Il peut être nécessaire de revenir maintenant sur l'un des aspects fondamentaux associé à la notion de structure préalable, à savoir quelle est la nature des informations qui nous permettent - comme dans le cas des figures 3.4.b et 3.5.b - d'organiser l'espace de sorte à pouvoir en isoler verbalement une partie à l'aide d'un GN démonstratif ou d'une description définie spatiale. La configuration unique que nous avons adoptée jusqu'ici reposait sur une répartition spatiale d'objets par ailleurs en tous points similaires. Cette répartition fait apparaître de fait une *frontière* qui est superposable à la discrimination spatiale exprimée par l'expression référentielle (orientée dans le cas de la description définie spatiale, isotrope dans le cas du démonstratif). L'hypothèse que nous souhaitons soutenir ici est qu'une telle répartition peut s'opérer le long de toute propriété des objets considérés, pour peu que celle-ci soit de nature perceptive, ou ait des conséquences perceptives (dans le cas par exemple d'une distinction fonctionnelle qui puisse être observée), et qu'elle permette d'établir une séparation claire entre deux zones de l'espace dans lesquelles la propriété prend deux valeurs différentes pour les objets considérés. La figure 3.10 illustre le cas de la propriété de couleur qui permet d'opérer le contraste attendu, mais on pourrait tout aussi bien avoir des différences de taille (deux petits, un grand), de texture, etc.

⁷ Dans le cas des deux exemples 3.7 et 3.8, il reste la possibilité d'un rôle *qualificatif* et non plus *déterminatif* associé au complément de nom, qui inactive celui-ci pour la recherche référentielle. Le résultat de l'interprétation serait alors dans les deux cas le groupe complet des triangles (seule la discrimination triangle-cercle est activée).



Ex. 3.10 : les triangles de gauche

Il se peut malgré tout (et c'est naturel si on considère que le locuteur a en quelque sorte une intention d'avance sur le système de dialogue) que le locuteur ne s'appuie pas sur une structure de discrimination directement accessible au système pour construire son expression référentielle. En effet, le locuteur peut très bien s'appuyer sur des informations sous-jacentes (lien fonctionnel entre les objets) qui ne sont pas directement visibles, ou encore projeter sur l'espace de visualisation un découpage lié à l'action qu'il va entreprendre (par exemple mettre en noir des triangles pour aboutir à une situation telle que celle observable figure 3.10 ci-dessus). Dans ce cas, il a bien sûr la possibilité d'affiner sa référence, en la sur-spécifiant par une contrainte de nombre (*les deux triangles de gauche*), mais il peut tout aussi bien « oublier » d'être explicite et se contenter d'une expression telle que celle que nous avons rejetée figure 3.4.a. Il est alors indispensable de mettre en place des mécanismes de rattrapage qui assouplissent les règles strictes que nous avons énoncées. Un tel rattrapage va consister à supposer l'existence de la structure la plus simple possible (afin de ne pas imposer trop de contraintes à la suite du dialogue) qui permette d'interpréter l'expression référentielle. Ce type de mécanisme est typique de nombreuses situations de dialogues homme-homme où l'on justifie *a posteriori* une référence faite par son interlocuteur en projetant une nouvelle structure sur l'univers observé⁸.

Un premier bilan - Nous avons justifié que la référence devait nécessairement s'appuyer sur une discrimination préalable du référent. Il est clair qu'un locuteur ne peut faire identifier à son interlocuteur que des objets isolables (et donc isolés d'une certaine façon). Mais, au delà même de cette contrainte de bon sens, l'expression langagière à elle seule ne peut créer la structure nécessaire. Ainsi dans l'exemple 3.4a, l'expression *les triangles de gauche* bien qu'elle soit "logiquement" suffisante par sa contrainte de nombre est inappropriée. Nous avons brièvement posé les contraintes imposées par les différents types d'expressions :

- une rupture catégorielle doit se superposer à la rupture perceptive dans le cas du défini simple
- une rupture de saillance doit se superposer à la rupture perceptive dans le cas du démonstratif
- une rupture spatiale doit recouvrir la rupture perceptive dans le cas du défini statial.

Reste maintenant à limiter l'extension de la rupture considérée. Qu'il s'agisse en effet de phénomènes de focus dans le dialogue ou de la simple extension du champ visuel dans le cas perceptif, on observe, dans les corpus de dialogue, des expressions qui ne se justifient que si on ne prend en considération qu'un sous-ensemble des objets représentés. Par ailleurs, comme nous le verrons, dans de tels sous-ensembles, des contraintes fonctionnelles peuvent jouer et définir en particulier des systèmes d'orientation spécifiques.

4. Effets de localisation

Dans ce qui précède, nous nous sommes limité à des exemples minimaux à base de figures géométriques. Dès lors, et pour autant que l'ensemble des objets représentés puisse être

⁸ *Ouah, tu as vu la voiture !* (au milieu d'une rue pleine de véhicules, mais où semble-t-il seule la Ferrari flambant neuve est digne d'être considérée comme une voiture...)

appréhendé globalement, le contexte d'interprétation des expressions était cet ensemble des objets présents. Par ailleurs, aucun lien fonctionnel n'apparaissait entre les objets.

Dans la suite, nous tirerons nos exemples d'une tâche d'aménagement d'appartements qui a servi de support à une expérimentation de type magicien d'Oz (Mignot 1993 et 1995) : des sujets devaient suivre des consignes du type *aménagez votre salon en veillant particulièrement à votre confort*. Pour cela, ils croyaient utiliser un système de dialogue permettant de s'exprimer par oral et d'utiliser des gestes (en fait le système en question était simulé par deux compères).

L'écran devant lequel étaient les sujets représentait un salon vide à côté duquel étaient représentés des meubles (cf. figure 4.1).

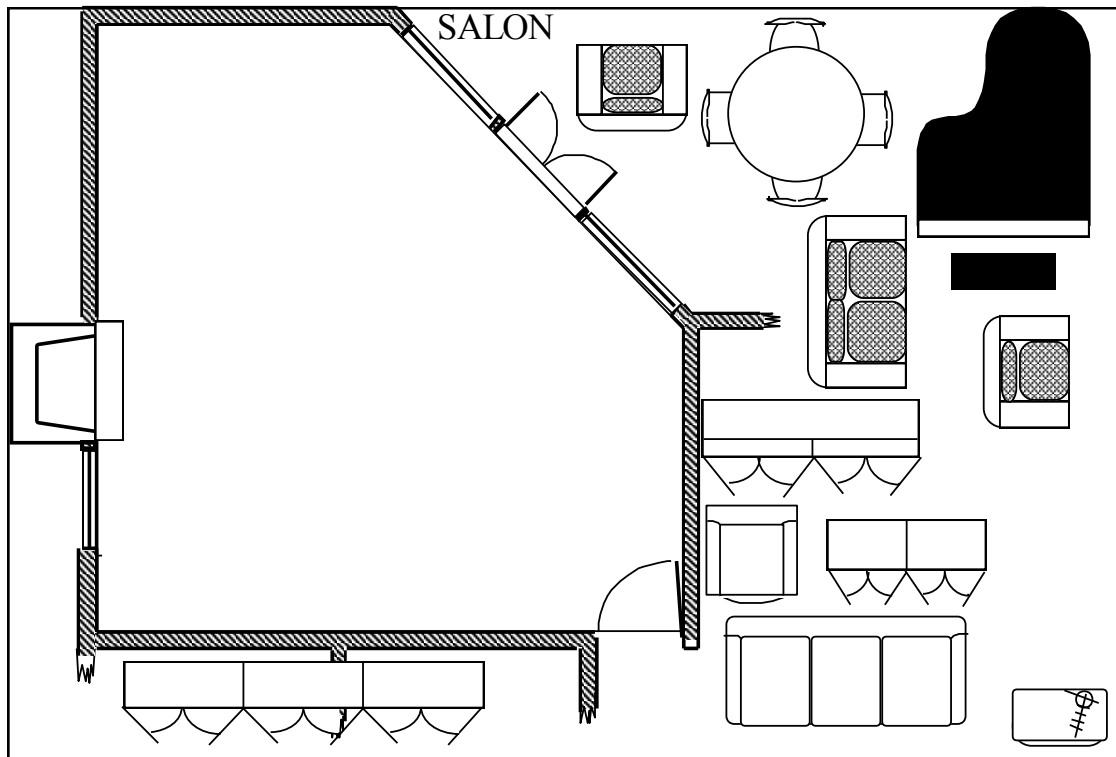


Figure 4.1 : Scène initiale du scénario d'aménagement intérieur.

Dans un tel contexte, deux *cadres* pour l'interprétation des expressions référentielles apparaissent immédiatement :

- le salon à l'intérieur duquel on dispose les meubles ;
- la réserve de meubles extérieure au salon.

4.1 Zones d'ancrage des structures

Dès lors que l'on a constaté que l'interprétation des expressions référentielles en contexte spatial nécessitait la mise en évidence d'une structure préalable au sein d'un groupe d'objets, il apparaît très vite que cette structure ne s'exprime que dans le cadre de contextes localisés qui en limitent l'extension, tout en en définissant les principales caractéristiques (type des objets présents, contraintes fonctionnelles, etc.). Ce phénomène est illustré par l'exemple de dialogue suivant, associé à l'environnement graphique de la figure 4.2.

U1 : *Retourne le divan*

U2 : *Remonte un peu le fauteuil de droite*

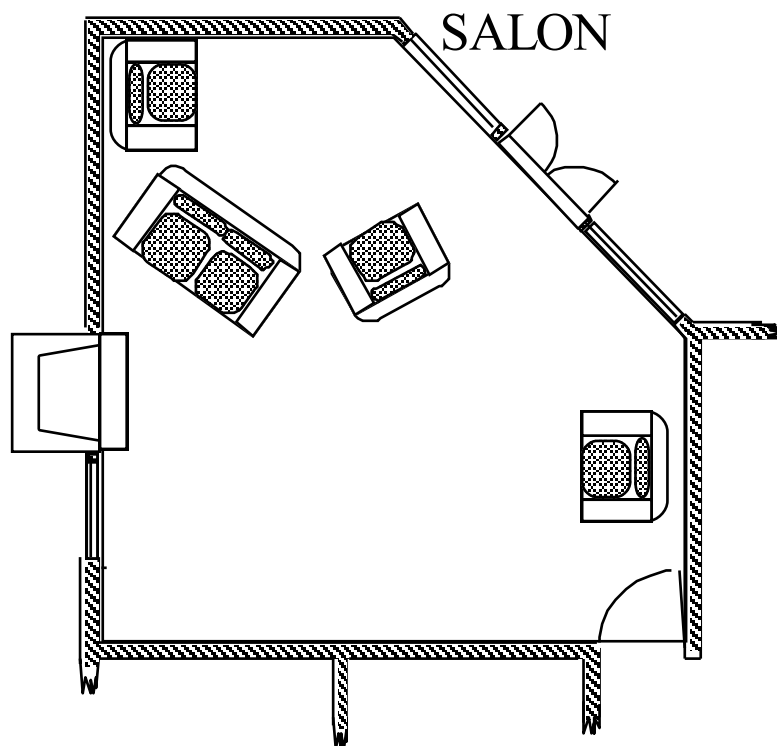


Figure 4.2 : Contextualisation référentielle.

Clairement, il semble difficile de projeter une structure spatiale globalement sur un espace de visualisation. Typiquement, en reprenant l'exemple ci-dessus, l'expression *le fauteuil de droite* n'est pas interprétée sur l'ensemble de l'écran mais bien dans un sous-ensemble lié au divan. Un tel sous-ensemble se caractérise non seulement par son inclusion dans le salon (qui exclurait en particulier les fauteuils présents dans les meubles en vrac) mais aussi par la nature des liens fonctionnels qui s'établissent entre les objets du sous-ensemble.

En particulier, au premier niveau de structure, les deux zones correspondant au salon et aux meubles en vrac respectivement sont systématiquement respectées dans le corpus (il n'a été observé aucune localisation d'un objet d'un des deux espaces par une référence spatiale relative à un objet de l'autre espace). Par ailleurs, aucune référence spatiale n'est jamais utilisée pour faire référence aux objets du vrac de meubles. Ce qui pourrait paraître surprenant au premier abord, s'explique bien par la différence fonctionnelle entre les deux espaces de localisation des meubles, puisque seul l'espace du salon actualise les possibles relations fonctionnelles entre les différents objets⁹.

Par ailleurs, si on considère maintenant le contexte englobant, la projection d'une instruction référentielle sur celui-ci filtre de nombreuses caractéristiques qui ne sont pas pertinentes pour le calcul référentiel. Il s'agit d'une part de toutes les parties du contexte qui ne sont pas susceptibles d'entrer dans le jeu d'alternatives en lien avec la relation fonctionnelle que nous avons mentionné. Le contexte est de fait rattaché temporairement (au cours des quelques échanges relatifs au même point de vue sur la tâche) à un contexte abstrait induit par la résolution de l'expression référentielle. Ceci apparaît sur des exemples tels que :

⁹ Lorsqu'il est nécessaire de désambigüiser une référence au sein du vrac, seuls des critères visuel de type couleur ou taille sont utilisés (S: *Quel divan ?* U: *Le plus petit.*), ceci pouvant être la plupart du temps interprété en termes d'intentions de la part de l'utilisateur.

U1 : Mets le divan devant la cheminée.
 S1 : Lequel ?
 U2 : Celui-ci (en montrant le divan gris)
 U3 : Mets les fauteuils de chaque côté.
 S2 : OK

où, alors même que les fauteuils pourraient désigner les quatre fauteuils présents (deux blancs et deux gris) la résolution de l'expression le divan (après négociation) livre son contexte d'interprétation à savoir le cadre : {divan gris, fauteuil gris 1, fauteuil gris 2}.

4.2 Caractérisation des contraintes de localisation

Nous pouvons à ce stade caractériser une opération de référence spatiale par le jeu d'un double système de relations verticales et horizontales au sein d'un contexte englobant l'objet référencé ainsi qu'un certain nombre d'alternatives implicitement ou explicitement posées au cours de l'acte de référence. La figure 4.3 résume ces contraintes pour une référence à un objet O1 dans un contexte C, pour un ensemble d'alternatives schématisé par l'objet unique O2.

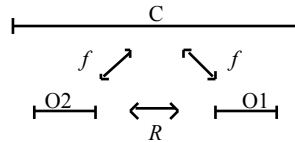


Figure 4.3 : Liens de localisation

Dans ce schéma, f est le lien fonctionnel (LF) qui unit les objets à leur contexte et qui doit être partagé par l'ensemble des alternatives. R est la relation de discrimination qui permet d'isoler l'objet référencé des autres membres de l'ensemble des alternatives. L'existence de C et de f est supposée être impliquée par l'expression référentielle ($LS(O1, O2, R) \Rightarrow \exists C \exists f (LF(O1, C, f) \wedge LF(O2, C, f))$). Pour les trois types d'expressions que nous avons observées, la relation R peut être spécifiée de la manière suivante :

- description définie (le triangle) : contraste inter-catégoriel ($Cat(O1) \neq Cat(O2)$)
- groupe nominal démonstratif : contraste intra-catégoriel ($Cat(O1) = Cat(O2)$)
- description définie spatiale : contraste de localisation ($Pos(O1) \cap Pos(O2) = \emptyset$)

Dans ce dernier cas, l'expression référentielle explicite plus finement la relation de contraste et rend donc la présence des alternatives encore plus sensible. Pour les deux autres situations, nous avons vu que les phénomènes de reprise ultérieure dans un dialogue permettaient de justifier l'hypothèse que nous avons faite de la présence implicite de ces mêmes alternatives.

Une conséquence immédiate d'une telle contrainte est qu'il est impossible de mettre en relation spatiale deux objets rattachés à des cadres de référence disjoints, à moins bien sûr de faire explicitement référence à la relation (d'ordre supérieur) qui existe entre ces deux cadres ou de considérer un autre point de vue qui permette de relier ces deux objets dans un même cadre. Ainsi, il sera difficile de dire de Pierre qu'il est à droite de Jean si Pierre et Jean sont dans deux bureaux différents du même immeuble. On peut malgré tout envisager soit de situer l'un des bureaux par rapport à l'autre (et par « transitivité » de situer l'un des personnages par rapport à l'autre), soit d'adopter un autre système de repérage, comme dans *Pierre est à l'autre bout de*

l'immeuble par rapport à Jean, faisant ainsi explicitement référence à un contexte englobant unique.

Une autre conséquence de la contrainte de cadre associée à l'interprétation des expressions spatiales est que les contextes ainsi construits sont stables pour les énoncés ultérieurs. Par exemple, dans la situation de la figure 4.4 associée à l'instruction suivante :

U : *Remonte un peu le fauteuil de droite et descend un peu celui de gauche*

Il est tout simplement impossible que la deuxième expression soit recalculée à un autre niveau que celui où la première a été interprétée, que ce soit au niveau de l'écran tout entier (en supposant qu'il s'y trouve d'autres fauteuils), ou même au niveau du salon. De telles observations permettent en particulier d'envisager de contraindre plus fortement l'interprétation d'expressions en 'celui', au delà d'une simple reprise d'une tête nominale¹⁰.

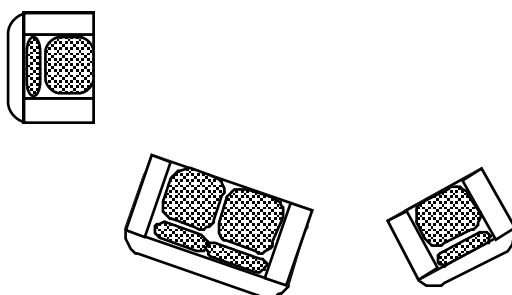


Figure 4.4 : *Préservation des contextes d'interprétation.*

Il reste malgré tout que des cas flagrants d'ambiguïté existent, en particulier quand les contextes de localisation (ou cadres de référence, cf. (Schang et Romary 1994, Schang 1995)) emboîtés ne correspondent pas à des distinctions fonctionnelles aussi marquées. C'est le cas par exemple de la séquence d'instructions suivante, associée à la mise en place de plusieurs fauteuils.

U1 : *Remplace les fauteuils de gauche par des fauteuils blancs*

U2 : *Remonte un peu celui de droite*

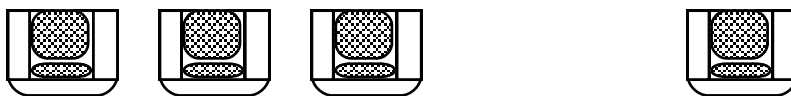


Figure 4.5 : *Un cas d'ambiguïté contextuelle.*

Dans ce cas, il y a une réelle ambiguïté référentielle lors de l'interprétation du deuxième énoncé, provenant du fait que la première intervention introduit un nouveau contexte de localisation qui interfère avec le contexte visuel préalable pris globalement. La situation est schématisée figure 4.6, où C1 représente le contexte visuel initial dans lequel l'expression *les fauteuils de gauche* permet de contraster le groupe d'objets O1 (composé de deux fauteuils) par rapport à O2. Une fois mentionné, O1 peut alors très bien être vu comme un nouveau contexte

¹⁰ Bien qu'une littérature abondante existe sur l'analyse linguistique des expressions référentielles construite à partir de 'celui' (Corblin 1990, Kleiber 1994), il n'existe pas à notre connaissance de véritable travail portant sur leur possible interprétation automatique.

actif C2 qui dans ce cas relève exactement du même type de contexte abstrait que C1. L'expression *celui de droite* peut alors s'ancrer soit sur C1 pour isoler O2, soit sur C2 pour isoler O13. D'un point de vue dialogique, la conséquence d'une telle analyse est que le système qui détecte une ambiguïté de ce type pourra réagir de manière relativement fine, sans réinterroger globalement le locuteur sur l'objet qu'il souhaitait désigner, mais plus sur le contexte exact qu'il sous-entendait. Par exemple (en reprenant la situation de la figure 4.5) :

S : *Le fauteuil complètement à droite ?*

U : *Oui, oui.*

ou

U' : *Non, l'autre.*

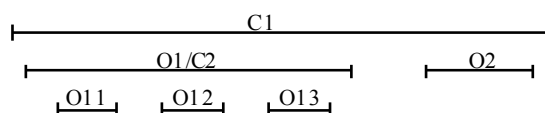


Figure 4.6 : Schématisation de l'ambiguïté sur les contextes imbriqués

Il est clair par ailleurs que le locuteur a toujours la possibilité d'utiliser des descriptions non ambiguës pour désigner les référents sur lesquels il souhaite travailler. Ainsi, O2 peut être isolé dans C1 par une reprise à l'aide de *l'autre fauteuil* et O13 dans C2 à l'aide de *le plus à droite des trois* par exemple¹¹. Il n'en demeure pas moins que dans un dialogue, le locuteur n'a pas toujours conscience qu'il ne travaille pas sur les mêmes contextes que son interlocuteur. Simon Garrod mentionne ainsi le cas d'une ambiguïté persistante apparue dans un dialogue recueilli lors d'une expérimentation d'interaction homme-homme portant sur une tâche de parcours de labyrinthe (Garrod et Anderson 1987, Garrod et Doherty 1994). Le contexte visuel est celui de la figure 4.7 associé à l'expression *la deuxième boîte*. Dans ce cas, il s'agit bien de la projection de deux schémas différents sur le même espace visuel, correspondant à des choix perceptifs différents de la part du locuteur et de son partenaire de jeu.

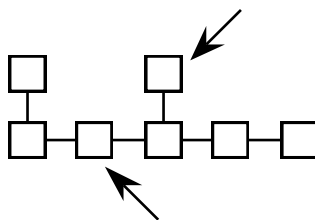


Figure 4.7 : un cas de disjonction contextuelle.

On remarque au passage que dans notre exemple initial (figure 4.5), tout comme dans l'exemple du labyrinthe figure 4.7, ni le contexte C1 ni le contexte C2 ne correspondent à des objets bien délimités. Les contextes d'ancrage (ou cadres) portent beaucoup d'information, mais pas nécessairement la catégorie de leur support. On peut se demander d'ailleurs si cette

¹¹ Par ailleurs, le rôle du prédicat peut devenir très important. Comparer les deux variantes de U2 :

U2' : *Mets celui de droite en vert.*

U2'' : *Eclaircis légèrement celui de droite.*

De façon non surprenante, il y a recouvrement des contextes référentiels avec les contextes intentionnels dans un dialogue.

catégorie peut réellement être définie en dehors d'un cadre englobant plus large ? Il ne nous semble pas que cela soit possible. Sans chercher à démontrer cette hypothèse, on remarque que dans l'exemple générique d'interface graphique pour un logiciel de dessin, il n'est pertinent de parler de l'écran que si celui-ci peut être intégré dans un jeu d'alternatives (le clavier, la caméra etc.) où différentes fonctions peuvent être réalisées (par exemple *éclaircir l'écran*). Le point de vue projeté est alors plus large que l'écran puisqu'il a trait aux éléments d'interaction entre l'homme et la machine.

A titre d'illustration, le déterminement d'une granularité associée à la subordination de toute opération de référence à un cadre spatial apparaît de façon flagrante lorsque l'acte de référence est associé à une prédication de positionnement. Ainsi, si on considère l'instruction suivante :

U : *Mets le piano dans le salon.*

la description définie *le salon* s'interprète dans un cadre correspondant à l'ensemble des pièces de l'appartement (ou, exprimé d'une autre façon, comme l'écran vu sous l'angle d'un ensemble de pièces). Dès lors, bien que l'instruction donnée consiste à insérer un meuble dans une pièce, le niveau de granularité considéré est celui de la pièce et non pas de celui des meubles au sein d'une pièce. Ainsi, l'intention exprimée par le locuteur de l'énoncé ci-dessus sera réalisée à partir du moment où le piano apparaîtra *quelque part* dans la pièce considérée, pour peu que le système se conforme, dans l'exécution de son action, aux contraintes générales qui peuvent régir l'insertion d'un meuble dans une pièce^{12,13}

5. Perspectives

Dans ce chapitre, nous avons essayé de dresser un bilan des différentes expressions référentielles susceptibles d'être rencontrées dans le cadre d'un dialogue homme-machine qui ferait intervenir une forte composante graphique. Plus que de chercher à décrire ces expressions, nous avons voulu faire ressortir la nature profonde des instructions d'interprétation que celles-ci semblent exprimer pour être en mesure à terme de mettre en œuvre de la façon la plus unifiée possible les traitements correspondant au niveau d'un système de dialogue. Bien que la description obtenue soit loin d'être exhaustive, il nous a semblé qu'un certain nombre de facteurs liés à la localisation de l'interprétation dans un cadre structuré pouvaient avoir une portée relativement large, indépendamment d'une tâche particulière qu'il s'agirait concrètement d'interfacer. Dès lors, les études complémentaires à réaliser devront faire

¹²Typiquement, le système devra proposer de mettre un poste de télévision à côté d'une prise de courant, etc.

¹³ Ce type de contraintes existe par exemple dans le cas de l'interaction avec un annuaire électronique. L'interprétation d'un énoncé tel que :

Ajoute ce numéro dans mon annuaire personnel.

conduira à l'inscription du numéro désigné dans le champ réservé aux numéros de téléphone et non pas dans un champ dédié au nom ou à l'adresse par exemple. Cette contrainte n'est cependant pas explicite dans l'instruction du locuteur qui souhaite voir apparaître le numéro dans un annuaire *particulier* par opposition par exemple à son annuaire d'entreprise. La détermination du niveau précis d'intentionnalité d'un locuteur, à l'aide notamment d'une interprétation fine des expressions référentielles, est une condition essentielle pour aboutir à des dialogues qui ne seront pas constamment interrompus par des demandes de précision de la part du système (par exemple : *Où exactement ?*).

apparaître de façon plus nette — même si nous avons abordé ce problème de façon occasionnelle ici — la nature des contraintes qu'une tâche telle que l'ATC (Contrôle du Trafic Aérien) fait peser sur ces mécanismes et en conséquence, la nature des informations, notamment lexicales, qu'il faut intégrer à nos systèmes.

Par ailleurs, nous avons abordé le problème de l'accès référentiel à un espace graphique sous l'angle quasi exclusif des expressions référentielles, au risque de négliger certains phénomènes propres au geste et qui pourraient, si ce n'est remettre en cause, au moins modifier les hypothèses que nous avons avancées. D'autres études (Bellalem et Romary 1995b, Bellalem et Romary 1995a) ont effectivement montré qu'une étude fine des trajectoires gestuelles permet de mettre en évidence une complémentarité importante entre le contenu instructionnel des expressions référentielles démonstratives par exemple et les segments significatifs d'un geste. De plus, il apparaît là encore que la perception, ainsi que les contraintes liées à la tâche jouent un rôle très important pour interpréter des mouvements qui sans cela pourraient avoir de multiples significations. Ainsi, le geste peut participer à l'opération de discrimination que nous avons mis en avant dans le cadre de l'interprétation des expressions référentielles. En effet, deux cas peuvent se présenter. D'une part, il se peut que la discrimination exprimée par le locuteur repose sur un ensemble d'objets dont la structure interne est directement perceptible par l'interlocuteur. Le geste n'aura alors qu'à pointer sur l'un des éléments de cet ensemble, ou éventuellement s'orienter de manière à exprimer un contraste net entre l'élément et les alternatives dans l'ensemble, comme dans l'exemple 5.1 (*ce triangle*).



Figure 5.1 : désignation d'un élément isolé.

Un cas de même nature peut correspondre à un groupe d'objets possédant une cohésion perceptive telle qu'il peut être considéré comme un tout et donc être pointé par ce que nous appelons une *désignation centrale*. Ainsi, dans la situation de la figure 5.2, l'expression *ces triangles* est parfaitement interprétable.

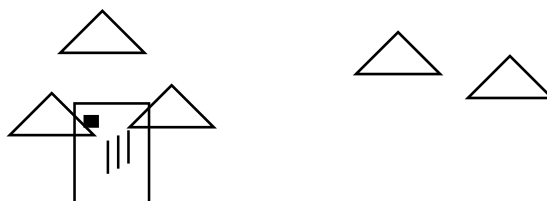


Figure 5.2 : désignation d'un groupe d'objets cohésifs.

Au contraire, lorsque la structure projetée par le locuteur sur l'espace graphique n'est pas implicitement accessible à l'interlocuteur (le système de dialogue en l'occurrence), le geste peut très bien servir à rendre celle-ci explicite. La trajectoire gestuelle servira alors à marquer les frontières invisibles dans le cadre de ce que nous appelons une *désignation périphérique*. C'est le cas de la figure 5.3 associée à l'énoncé *ces deux triangles*.

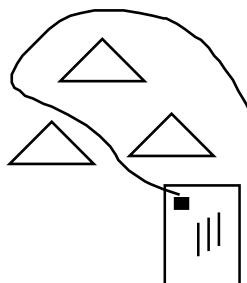


Figure 5.3 : exemple de désignation périphérique.

On le voit, la mise en place d'une composante référentielle précise dans un système de dialogue homme-machine requiert des études fines tant dans le domaine de la langue que dans le domaine du geste. Ces études, même si elles doivent se nourrir des travaux existant en linguistique notamment, doivent prendre en compte les contraintes spécifiques de la mise en œuvre de dialogues finalisés, en particulier en ce qui concerne la gestion des intentions de l'utilisateur. Le travail pluridisciplinaire correspondant, où la psychologie cognitive doit occuper une place importante, passe par une bonne connaissance réciproque des préoccupations de chaque domaine, sans pour autant bien sûr qu'aucune des disciplines concernées n'ait une quelconque priorité du point de vue de ses méthodes ou de ses modèles.

7. Références

- Bellalem Nadia et Laurent Romary 1995a, Langue et geste pour le dialogue homme-machine finalisé, in *Communication en conception*, K. Zreik et J. Caelen (Ed.), EuropIA, p. 185-201.
- Bellalem Nadia et Laurent Romary 1995b, Reference interpretation in a multimodal environment combining speech and gesture, Actes First International Workshop on Intelligence and Multimodality in Multimedia Interfaces, Edinburgh.
- Cadoz Claude 1992, Le geste canal de communication homme/machine - la communication instrumentale, *Technique et Science Informatique*, 13, 1, p. 31-61.
- Corblin F. 1987, *Indéfini, défini et démonstratif*, Droz, Genève.
- Corblin Francis 1990, Les groupes nominaux sans nom du français, in *L'anaphore et ses domaines*, G. Kleiber et J.-E. Tyvaert (Ed.), Klincksieck, Paris, p. 63-80.
- Dale Robert 1995, Generating One-Anaphoric Expressions: Where Does the Decision Lie?, Working Papers of PACLING-II, Brisbane, Australia, p. 49-58.
- Dauchy P., C. Mignot et C. Valot 1993, Joint speech and gesture analysis : some experimental results, Actes Eurospeech 93, p. 1315-1318.
- Falzon Pierre 1986, *Langages opératifs et compréhension opérative*, Thèse de Doctorat, Université Paris V.
- Gaiffe Bertrand, Jean-Marie Pierrel et Laurent Romary 1992, De la référence à la modélisation de la tâche : vers une certaine idée du dialogue homme-machine, Actes Séminaire Dialogue des Pôles Parole et Langage Naturel (GDR-PRC CHM), Dourdan, France, P. 117-129.

- Garrod S. et A. Anderson 1987, *Saying what you mean in dialogue: a study in conceptual and semantic co-ordination*, *Cognition*, 27, p. 181-218.
- Garrod Simon et Gwyneth Doherty 1994, *Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions*, *Cognition*, 53, p. 181-215.
- Kleiber Georges 1994, *Anaphores et pronoms*, Editions Duculot, Louvain-la-Neuve (Belgique).
- Mignot C., C. Valot et N. Carbonell 1993, *An Experimental Study of Future "Natural" Multimodal Human-Computer Interaction*, Actes INTERCHI'93 1993 Conference on Human Factors in Computing Science INTERACT'93 and CHI'93, Amsterdam (The Netherlands).
- Mignot C. 1995, *Usage de la parole et du geste dans les interfaces multimodales - Etude expérimentale et modélisation*, Doctorat d'Université, Université Henri Poincaré, Nancy I..
- Oskan Nadine 1994, *Analyse de dialogues finalisés dans une perspective communicationnelle*, Thèse de Doctorat en Sciences Cognitives, Institut National Polytechnique de Grenoble.
- Pouteau X., L. Romary et J.-M. Pierrel 1994, *Voix, geste et multimodalité : quand dire c'est faire faire*, Actes Congrès ERGO-IA'94, IDLS, Biarritz, p. 491--500.
- Reiter Ehud et Robert Dale 1992, *A Fast Algorithm for the Generation of Referring Expressions*, Actes Proceedings of the 15th International Conference on Computational Linguistics, volume I, Nantes, France, p. 232-238.
- Romary L. 1993, *L'interprétation de ici dans des énoncés de positionnement*, Actes Workshop « Le dialogue homme-robot en langage naturel : problèmes psychologiques », Presses Universitaires de Caen, Université de Caen.
- Romary L., B. Gaiffe et J. Pierrel 1993, *Multimodality: why and how?*, Actes du Workshop ERCIM on Multimodal Human-Computer Interaction, Nancy, p. 59-62.
- Schang D. 1995, *Application de la notion de cadre aux énoncés de positionnement et de référence*, Centre de Recherche en Informatique de Nancy, n° 95-R-048.
- Schang D. et L. Romary 1994, *Framing the world, towards a localised spatial reasoning*, Actes 3rd International Conference on the Cognitive Science of Natural Language Processing (CSNLP-94), Dublin.
- Souvay G. et J.-M. Pierrel 1995, *DIAPASON A Development Environment for the Integration of an Oral Input in Machine Control Applications*, Proceedings of ESCA Tutorial and Workshop on Spoken Dialogue Systems - Theories and Applications, Vigso (Denmark), p. 133-136.
- Sperber Dan et Deidre Wilson 1986, *Relevance, communication and cognition*, Basil Blackwell, Oxford.
- Vandeloise Claude 1986, *L'espace en français*, Editions du Seuil, Paris.