



HAL
open science

Langue et geste pour le dialogue homme-machine finalisé

Nadia Bellalem, Laurent Romary

► **To cite this version:**

Nadia Bellalem, Laurent Romary. Langue et geste pour le dialogue homme-machine finalisé. Caelen, Jean and Zreik, Khaldoun. Le Communicationnel pour Concevoir, Europa Productions, pp.185-201, 1995. hal-00416585

HAL Id: hal-00416585

<https://hal.science/hal-00416585>

Submitted on 14 Sep 2009

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 et geste pour le dialogue homme-machine finalisé

Nadia Bellalem & Laurent Romary

CRIN-CNRS & INRIA Lorraine

RÉSUMÉ. Nous présentons, dans cet article, une étude visant à intégrer le geste dans la communication homme-machine. Plus précisément, notre travail est centré sur le geste effectué dans le cadre de la référence aux objets (c'est le geste qui montre les objets dont on parle) ; ce dernier intervient donc en tant que complément d'une entrée langagière. Ainsi, nous montrons la nécessité d'une analyse approfondie du geste allant du signal issu du capteur de geste jusqu'à l'identification des objets visés, pour assurer la compréhension du message global parole+geste. Cette analyse se décompose en deux étapes. La première dite syntaxico-sémantique s'appuie sur la forme de la trajectoire pour repérer les parties significatives du signal auxquelles on associe une sémantique d'action de désignation. La seconde concerne l'interprétation du geste en contexte. Il s'agit là de prendre en compte d'une part l'application (notamment la scène visualisée) et ses caractéristiques ; d'autre part le dialogue oral et les caractéristiques instructionnelles de la référence langagière qui accompagne le geste analysé.

MOTS CLÉS. dialogue homme-machine, geste de désignation, reconnaissance du geste, interprétation du geste, déixis, référence aux objets.

1. Introduction

L'usage de la parole seule, à l'exclusion de tout autre mode de communication, peut sembler parfois inadapté lorsqu'il s'agit de piloter une application à caractère fortement visuel, dans le cadre d'une interface

graphique par exemple. En effet, l'accès aux objets par la langue impose de recourir à des expressions plus ou moins complexes. Dans ce cadre, le calcul de la référence fait l'objet de nombreux travaux, qui tentent d'identifier, en particulier, les mécanismes qui sous-tendent l'utilisation des différentes prépositions spatiales (à droite, sur, à côté, derrière, etc.) [Pribbenow, 93 ; Schang, Romary 94].

L'autre possibilité de faire référence aux objets du discours est d'utiliser, en association avec des expressions langagières à valeur déictiques, le geste de désignation qui présente l'avantage d'effectuer un accès direct et rapide aux objets [Wahlster, 91]. En fait, on se rapproche là de la situation de communication homme-homme où l'on montre du doigt les objets dont on parle.

Ces deux aspects de la référence aux objets sont complémentaires et doivent être pris en compte dans la conception de systèmes de dialogue, d'autant que des expériences de type magicien d'Oz [Mignot and al, 93] ont montré que les sujets y font largement appel.

Dans cette optique, nous nous intéressons au geste de désignation¹, considérant qu'il participe à la visée communicative au sein d'un dialogue homme-machine. Nous présentons à cet effet, une étude des indices linguistiques qui se prêtent à l'association avec le geste de désignation, puis une étude du geste explicitant les différentes étapes nécessaires à sa compréhension dans le cadre du dialogue.

2. Expressions langagières et désignation

2.1 Quelles expressions référentielles ?

¹ Nous ne considérons pas les autres usages du geste (ergotique, épistémique) tel que mis en évidence par [Cadoz, 94] par exemple.

La vision classique de la référence aux objets² permet de répertorier les marqueurs linguistiques susceptibles de prendre une valeur déictique pour construire des syntagmes nominaux démonstratifs [Cosnier and al, 82] :

- les adjectifs démonstratifs (*ce, cette, cet, ces*), éventuellement accompagnés d'une marque déictique (*ce N-ci, ce N-là*) ;
- le pronom démonstratif (*ça*) ;
- les déictiques purs notamment *ici* et *là*.

Il peut sembler étrange de ne retenir que ces marqueurs, alors que l'intuition laisse penser que le geste peut en accompagner un certain nombre d'autres. Sans entrer dans le détail d'une telle discussion, nous pouvons donner quelques éléments propres à alimenter la réflexion sur le sujet. En premier lieu, descriptions définies et indéfinies se rencontrent effectivement dans des situations de dialogue où un geste de désignation semble participer à leur interprétation. Il ne s'agit pourtant pas d'une réelle coréférentialité entre parole et geste, mais plutôt d'une focalisation sur un nouvel espace d'interprétation, au sein duquel le groupe nominal va pouvoir être interprété. L'un des arguments à l'appui de cette idée est par exemple la présence d'une indication préalable de focalisation concomitant au déclenchement du geste (*Là/regarde, un chat/le chat !*). Il est clair que dans le cadre d'une analyse contextuelle des expressions référentielles, nous ne pourrions nous permettre de négliger complètement ces cas, même s'ils ne semblent pas devoir être très fréquents dans le cadre d'un dialogue homme-machine. Un deuxième type d'expressions est plus facile à rejeter pour nous : les pronoms (e.g. *il/elle*). Ces expressions ne semblent pas pouvoir être accompagnées d'un geste, en dehors de leurs formes marquées (*lui/elle*) (Kleiber, 92). Même dans ce cas, la présence d'un sème du type /humain/ (*C'est lui le coupable* versus **C'est lui le fauteuil que je veux déplacer*) ne nous donne aucun remords à laisser ces expressions de côté.

2.2 Interprétation de ces expressions

² En situation de dialogue homme-machine finalisé, nous opposons cette notion à la référence aux actions. Le terme "objet" correspond aux entités manipulées dans le cadre d'une tâche particulière sans qu'à ce stade il ne soit nécessaire d'introduire une ontologie précise.

Pour pouvoir être interprétés, les indices linguistiques que nous avons retenus doivent être associés à un élément de focalisation qui, dans le cadre qui nous intéresse, peut être ramené au seul geste de désignation. Le syntagme nominal apporte lui une indication de nombre et de catégorie concernant les objets candidats à la référence. Dans un premier temps, nous pouvons faire les observations suivantes :

- les **groupes nominaux démonstratifs *ce N/cet N/cette N***, indiquent que la référence gestuelle porte sur un objet unique de type 'N',
exemple : *Ferme cette fenêtre* (plus un geste montrant la fenêtre)
- le **groupe nominal démonstratif *ces N***, indique que la référence gestuelle porte sur un groupe d'objets de même type 'N' en nombre indéfini,
- le **groupe nominal démonstratif *ces x N*** où x est un adjectif numéral, indique que la référence gestuelle porte sur x objets de type 'N',
- le **pronom démonstratif *ça*** indique simplement qu'il existe une référence gestuelle associée sans préciser le type ou le nombre d'objets visés,

- les **déictiques *ici et là*** indiquent que la référence porte sur un lieu.

Une étude plus fine de ces indices [Corblin, 87 ; Romary, 93 ; Gaiffe et al, 94] montrent qu'ils marquent en réalité l'insertion du référent dans un contexte réduit et structuré où celui-ci se différencie d'autres entités possédant des propriétés communes. Ce point sert de base à la définition d'une méthodologie pour l'étude du geste de désignation qui vise à mettre en évidence un certain nombre d'indices contrastifs.

Ainsi, le geste de désignation peut se définir comme étant le mouvement de la main dont le rôle est de guider le regard de l'interlocuteur vers une région particulière de l'espace visuel partagé. Son rôle est donc de focaliser l'attention de l'interlocuteur vers une région pour y isoler des objets ou un lieu du reste de la scène. Le canal verbal fournit un cadre pour l'interprétation de la référence. Il permet notamment dans le cas du geste de désignation de catégoriser les objets appartenant au sous-espace contrasté par le geste. C'est cette complémentarité qui permet de résoudre la référence langue+geste.

Afin de préciser cette vision contrastive de l'analyse co-référentielle d'un couple [expression référentielle + geste], regardons rapidement ce qui ce

passé dans le cas d'un groupe nominal démonstratif en "ce N". Nous avons montré [Gaiffe et al 94], à la suite de F. Corblin notamment [87], qu'une telle expression marque un contraste intra-catégoriel, c'est à dire où un élément de type 'N' se distingue d'autres éléments du même type de part sa présence en situation focale. Dans le cadre d'un accès à une présentation graphique de la tâche à l'aide d'une telle expression, ce principe d'opposition peut s'assimiler à un filtrage de l'espace de représentation de sorte que seuls des objets de type 'N' ne sont considérés, tout autre élément se trouvant opacifié à cette étape de l'interprétation. Dès lors, l'analyse de la trajectoire de désignation s'effectue sur la base du *niveau d'échelle* imposé par le filtrage en 'N'. En d'autres termes, la granularité recherchée pour l'interprétation du geste correspond au minimum nécessaire pour pouvoir différencier correctement un élément du type 'N' des autres éléments de la même catégorie. L'interprétation finale correspond alors à l'élément *le plus focal*, étant données les singularités (cf. infra - 3.2.1) rencontrées sur la trajectoire gestuelle.

Ainsi, dans le cadre de l'aménagement intérieur d'un appartement, la précision recherchée pour l'interprétation de [*cette pièce* + geste] (e.g. *mets un fauteuil dans cette pièce*) sera bien moindre que celle requise pour [*ce fauteuil* + geste] (e.g. *mets ce fauteuil en face de la cheminée*).

3. Analyse du geste de désignation

L'exploitation du geste dans le dialogue nécessite une étude approfondie allant de l'analyse du signal gestuel -issu du dispositif qui a servi à le capturer (souris, gant numérique, stylo, etc.)- jusqu'à l'identification des objets mis en évidence par la désignation. Nous avons déjà présenté [Bellalem & Romary 1993], l'étude menée pour extraire un certain nombre d'indices pertinents à partir des trajectoires obtenues d'une part à l'aide d'une souris³ et d'autre part à l'aide d'un gant de désignation, ce dernier

³ La souris est alors considérée comme ayant un comportement totalement neutre vis à vis de l'interface graphique. En particulier, tout déplacement de celle-ci est potentiellement interprétable comme un geste. Parallèlement, nous avons considéré dans le cadre des tests que nous avons conduits, que l'usage d'un bouton ne faisait qu'apporter des indices de segmentation pour l'analyse de la trajectoire.

dispositif posant de nombreux problèmes de mise en œuvre. Nous ne reviendrons pas ici sur cette étude, qui reposait sur un certain nombre de techniques classiques de reconnaissance des formes (approximation des trajectoires à l'aide de splines, critères de segmentation), mais, comme nous le verrons dans le paragraphe 3.2.2, nous pouvons mentionner qu'elle nous a fourni l'assurance d'avoir des informations d'une bonne fiabilité en préalable aux étapes d'analyse que nous présentons dans cet article.

De fait, l'objectif que nous nous fixons ici est de définir les moyens qu'il faut mettre en œuvre pour véritablement *comprendre* la trajectoire d'un geste, c'est à dire de pouvoir lui donner une signification relativement à la tâche de conception sur laquelle porte le dialogue. Il faut donc l'interpréter par rapport à la scène sur laquelle il intervient et par rapport au message verbal qui l'accompagne en tenant compte du contexte du dialogue, à savoir les éventuels effets de focalisation résultant des énoncés antérieurs du locuteur. Nous proposons dans ce qui suit de centrer notre analyse sur une tâche particulière qui d'une part semble bien se prêter à l'usage du geste de désignation et d'autre part présente des caractères génériques par rapport à l'activité de conception sur ordinateur. Nous essayerons par ailleurs d'explicitier les différentes étapes menant à la compréhension du message langue+geste.

3.1. Application choisie

Nous nous appuyerons sur une application de type "aménagement d'intérieur"⁴, pour laquelle la tâche consiste à manipuler des objets graphiques visualisés sur une scène. Les énoncés verbaux sont essentiellement des énoncés de positionnement d'objets qui ont la forme suivante⁵ :

4 Ce type de tâche est à rapprocher de l'expérience menée à l'EdCAAD à Edinburgh [Neilson & Lee 94] qui portait sur l'analyse de dialogues homme-homme multimodaux pour l'agencement d'une cuisine. Nous partageons par ailleurs un certain nombre de points de vue avec ces auteurs, notamment en ce qui concerne l'interaction entre langage et geste.

action sur un ensemble d'objets.

où l'ensemble d'objets est le résultat de l'interprétation d'un acte de référence.

Plus particulièrement, et afin de mettre en évidence les différents paramètres susceptibles d'intervenir dans l'interprétation du geste dans un tel cadre, nous centrerons notre analyse sur un exemple particulier de trajectoire de désignation, illustré par la figure 1⁶, en accompagnant du message verbal *Déplace ces deux fauteuils ici.*

⁵ Nous n'abordons pas en particulier tout ce qui touche à la gestion du dialogue, pour nous limiter exclusivement aux énoncés reliés à la tâche de conception, tout en sachant qu'une analyse des phénomènes référentiels dans un dialogue homme-machine ne peut occulter complètement cet aspect. A titre d'exemple, l'étude d'une suite d'énoncés telle que :

permet de mieux appréhender les phénomènes de contextualisation par exemple (cf. [Romary 94])

⁶ Schéma d'aménagement de salon emprunté à l'expérience Magicien d'Oz réalisée au CERMA en collaboration avec le CRIN [Mignot and al 93].

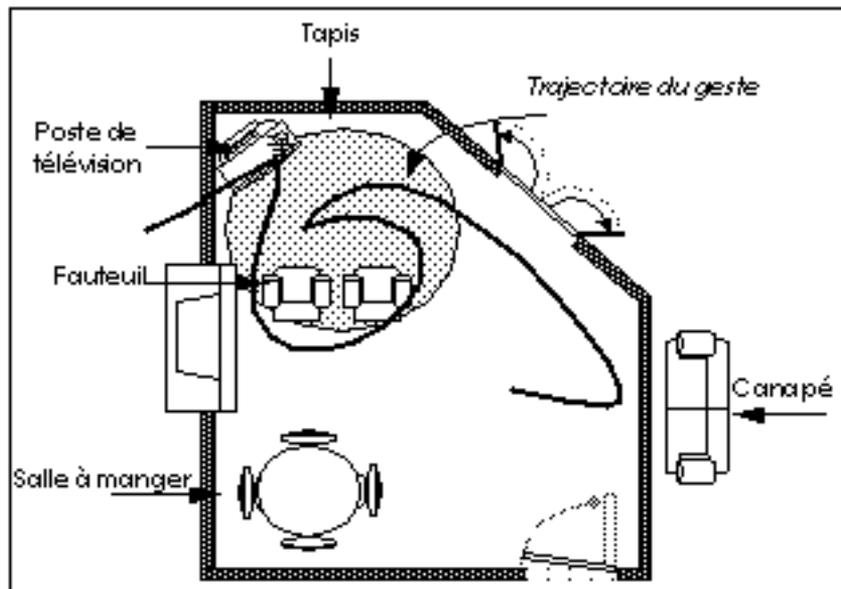


Figure 1 : Geste désignation pour une tâche d'aménagement de salon

3.2. Analyse syntaxico-sémantique du geste

La première étape de la compréhension du geste dans le cadre de la référence consiste à analyser le signal gestuel indépendamment de la scène sur laquelle il intervient, le but étant la détection des parties significatives du signal. Il faut en effet noter que le signal gestuel issu du capteur est un flux de données correspondant à tout le mouvement de la main, c'est-à-dire qu'il comprend la phase de rapprochement de l'écran, les désignations effectives relatives au message verbal émis, et la phase d'éloignement de la main⁷. Il est clair que seules les portions de signal correspondant aux désignations effectives sont importantes puisqu'elles concentrent toute la signification du geste. Le problème est de pouvoir les différencier du reste du signal. Par

⁷ Ceci dans le cas où l'on utilise un gant de désignation du type *Dataglove*. Les difficultés inhérentes à ce dispositif font que nos tests se sont portés essentiellement sur des trajectoires obtenues à l'aide d'une souris, sachant que nos résultats sont tout à fait concevables dans le cadre de dispositifs tels que des tablettes graphiques par exemple.

ailleurs, comme le geste de désignation a pour but de contraster un sous-espace de l'espace visuel, il doit comporter dans son expression des indices contrastifs. Ces derniers correspondent aux parties significatives du signal gestuel.

Dans ce contexte, nous proposons un modèle de ces indices contrastifs – que nous nommerons *singularités* – qui permet d'une part, de prendre en compte des formes diverses comme des désignations potentielles et d'autre part, qui assure la gestion d'éventuelles ambiguïtés.

3.2.1. Notion de singularité

La partie significative d'un geste est caractérisée :

- du point de vue de l'expression du geste, par une forme particulière que nous appelons *singularité* ;
- du point de vue du sens, par la possibilité de lui associer une sémantique d'action de désignation.

Une singularité s'observe relativement à une propriété de la trajectoire (par exemple la courbure), et relativement à un segment de trajectoire, stable pour la propriété considérée. C'est un événement local. L'examen complet d'un geste doit porter sur toutes ses propriétés et sur la totalité du signal ; la sémantique globale du geste est établie par association/regroupement des singularités détectées. En fait, la notion de singularité permet de mettre en relation la forme du geste et sa signification.

L'idée proposée est de construire une représentation symbolique de la forme du geste à partir de laquelle on recherchera les singularités en prenant comme hypothèse de travail, que toute singularité n'est pas accidentelle mais provient de l'intention de l'utilisateur de signifier quelque chose : en particulier, de désigner. Le cas des singularités réellement accidentelles (i.e. qui ne signifient rien pour l'utilisateur) seront éliminées au moment de l'interprétation du geste soit parce qu'aucune expression langagière ne les accompagne, soit parce qu'aucun objet ne peut leur être associé.

D'une façon générale, on peut distinguer trois types de singularités :

a - singularité ponctuelle :

Elles sont relatives à un point donné de la trajectoire décrite par le geste. Il s'agit d'un point marquant une rupture d'homogénéité avec son voisinage, c'est à dire qui présente une discontinuité sur au moins l'une des propriétés associées à la trajectoire.

Une singularité ponctuelle dans le geste a pour objectif de mettre en évidence un point particulier de l'espace de désignation. Par conséquent, la sémantique qu'on lui attribue est la désignation de l'objet ou du lieu en relation spatiale avec le point visé.

b - singularité de trajectoire simple :

Une singularité de trajectoire simple correspond à un segment de trajectoire non homogène avec les deux segments adjacents, relativement à un ensemble de propriétés. Ces derniers étant homogènes entre eux relativement au même ensemble de propriétés. Une boucle entourée de deux droites est un exemple de singularité de trajectoire relativement à la courbure.

c - singularité de trajectoire répétitive :

Ces singularités correspondent à une alternance de singularités ponctuelles homogènes entre elles par rapport à un ensemble de propriétés et de segments de trajectoire également homogènes entre eux. De l'agrégation de ces éléments résulte une singularité de trajectoire répétitive. Le zigzag effectué pour la désignation d'une région est une illustration de ce type de singularité.

D'une façon générale, une singularité est détectée sur un élément de trajectoire (point ou segment) relativement à une propriété P, si on observe un changement de la propriété P sur cet élément et une stabilité de P sur les éléments adjacents.

Cette définition est illustrée par le graphe temporel de la figure 2.

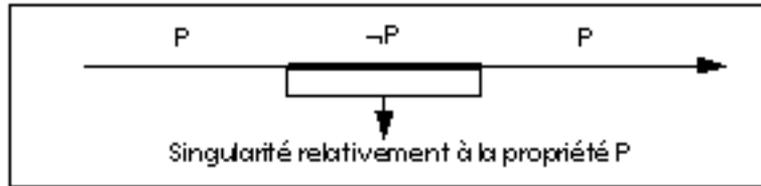


Figure 2 : Représentation temporelle des variations de la propriété P

3.2.2. Analyse syntaxique du geste

Il s'agit d'une modélisation de la trajectoire du geste c'est-à-dire de la construction d'une représentation de la forme du geste. Différentes propriétés permettent de construire cette modélisation. On peut citer :

a - la courbure :

L'étude du tracé de la courbure associée à chaque point de la trajectoire du geste permet de repérer les variations de la forme : une droite, une courbe, un point de rebroussement, un point d'inflexion.

b - la vitesse de déplacement pendant la production du geste :

Son étude permet d'une part, de cerner avec plus de précision les phases de rapprochement ou d'éloignement qui peuvent être plus rapides que les phases de désignations effectives ; d'autre part de détecter les points d'arrêt observés par l'utilisateur au moment de la production de son geste qui peuvent représenter une désignation ou correspondre à une hésitation.

c - les points de recouplement dans la trajectoire du geste :

Ces indications sont essentielles pour la détection des boucles présentes dans le geste qui, sans ambiguïtés, délimitent une région de la scène.

d - l'angle de courbure :

Il permet de rechercher les portions de trajectoire qui délimitent une région sans que celle-ci soit complètement entourée.

e - les indices spécifiques au capteur :

Il s'agit particulièrement des événements associés aux boutons de la souris tels que la pression, le relâchement d'un bouton. En effet, la présence d'un clic souris dans une trajectoire doit être considérée comme une

singularité ponctuelle dont le but est la mise en évidence d'un point particulier de la scène.

3.2.3. Analyse sémantique

L'examen des différentes propriétés de la trajectoire gestuelle a permis d'obtenir, pour un même geste, différentes singularités, et ceci en fonction des propriétés considérées. Se pose alors le problème de l'intégration de toutes ces connaissances pour proposer une sémantique globale du geste.

L'analyse sémantique consiste à reconnaître dans la modélisation du geste les singularités et à construire pour un geste la sémantique globale qui n'est autre que l'association des singularités.

Plusieurs cas sont à considérer :

1- Le cas des singularités disjointes pour lesquelles on n'observe aucun chevauchement entre des singularités de différents types ; le sens global est déterminé en considérant tour à tour les singularités : les singularités ponctuelles pour la désignation d'un point plus ou moins précis et les singularités de trajectoire pour la désignation de régions (voir figure 3).

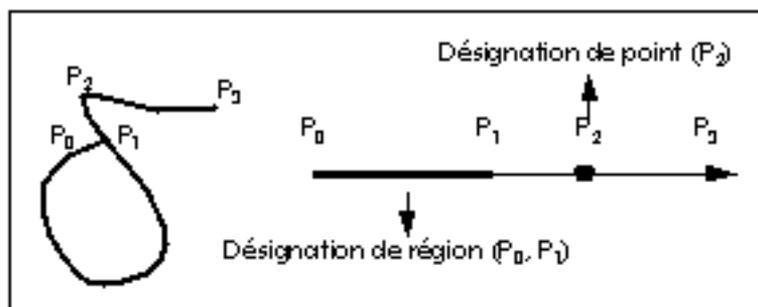


Figure 3 : Association de singularité disjointes

2- L'association d'une singularité ponctuelle et d'une singularité de trajectoire, dont un exemple pourrait être un point d'arrêt à l'intérieur d'une boucle. Il est difficile à ce stade de l'analyse de privilégier une hypothèse par rapport à l'autre. Ce cas peut être géré en conservant deux possibilités de sens pour le geste : la désignation de la région délimitée par la singularité

de trajectoire et la désignation du point correspondant à la singularité ponctuelle (voir figure 4).

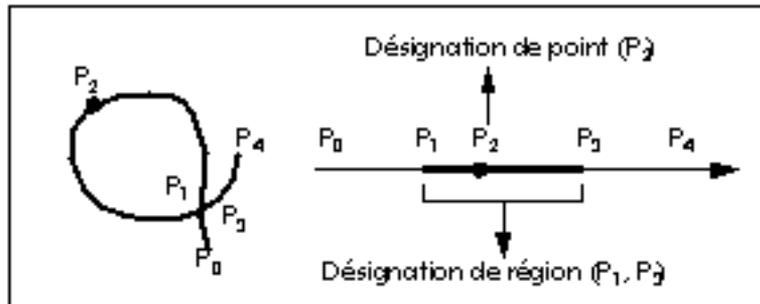


Figure 4 : Association d'une singularité ponctuelle et d'une singularité de trajectoire

3- L'association de deux singularités de trajectoire dont un exemple pourrait être l'existence d'une boucle dans le geste parallèlement à une pression continue sur un bouton de la souris. Là encore, on ne peut pas établir quelle singularité de trajectoire représente effectivement la désignation sauf s'il y a une juxtaposition parfaite entre elles. Dans ce dernier cas l'utilisateur, en maintenant le bouton de la souris enfoncé pendant l'entourage, veut mettre l'accent sur cette partie du geste (voir figure 5).

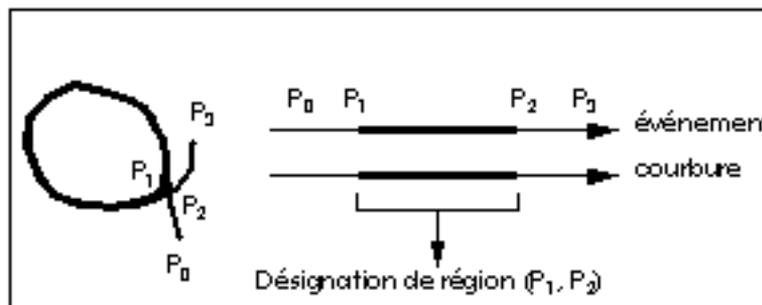


Figure 5 : Association de deux singularités de trajectoire

3.2.4. Un exemple

La modélisation syntaxique de notre exemple de geste permet de décrire la forme générale. Cette dernière s'exprime en terme élémentaire de forme : droite, courbe et point de rebroussement comme illustrée par la figure 6.

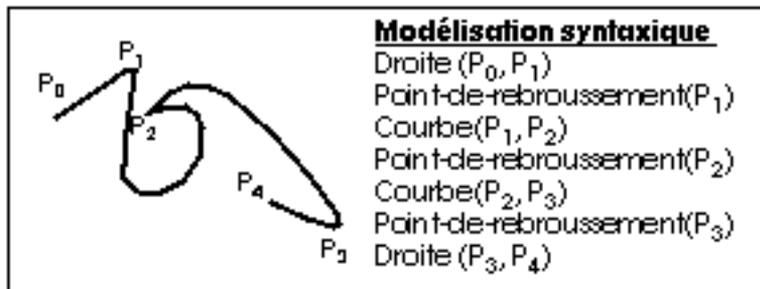


Figure 6 : Analyse syntaxique du geste de désignation

Ce qui permet de dégager, pour l'exemple proposé, deux hypothèses de sémantiques pour le geste effectué par association des singularités ponctuelles et des singularités de trajectoire. La première prend en compte les singularités ponctuelles survenant aux limites de la singularité de trajectoire, la seconde privilégie l'entourage semi-fermé (cf. figure 7). L'ambiguïté pourra être levée dans la phase d'interprétation.

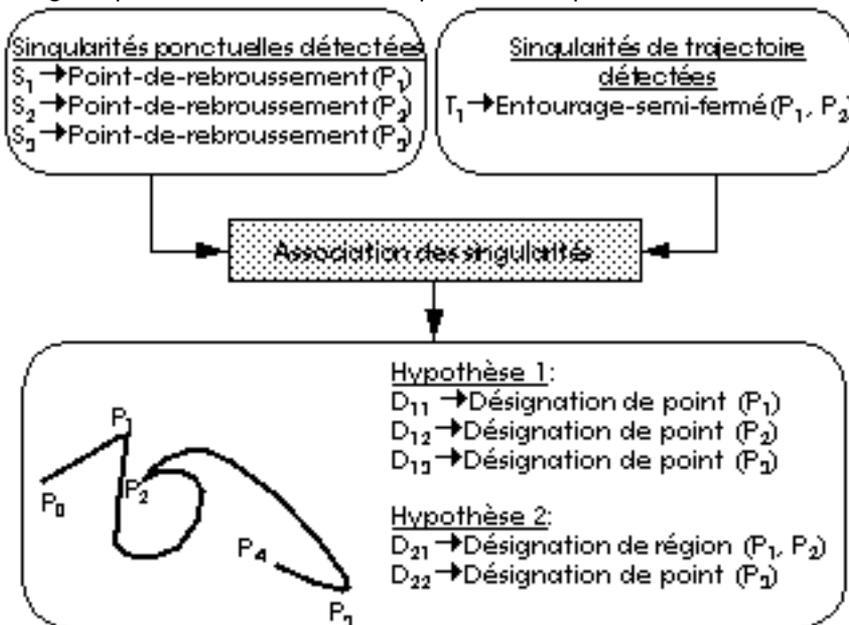


Figure 7 : Analyse sémantique du geste de désignation

3.3. Interprétation du geste en contexte

La seconde étape de compréhension du geste est l'interprétation dans le contexte où il a été produit. Ce qui signifie qu'il faut considérer d'une part les objets de la scène sur laquelle le geste intervient, d'autre part qu'il faut prendre en compte le message verbal qui l'accompagne.

3.3.1. Interprétation dans le contexte de l'application

Ce processus d'interprétation du geste en contexte nous conduit à modéliser l'espace visuel pour prendre en compte non seulement la répartition des objets dans l'espace mais, également, les relations entre les objets (spatiale, fonctionnelle) qui peuvent varier suivant les visualisations considérées (2D, 3D). Le but à atteindre est la construction d'une représentation se rapprochant de ce que perçoit un interlocuteur humain. Cette représentation doit prendre en compte les relations spatiales telles que la proximité ou encore la superposition des objets dans la scène auxquelles s'ajoutent les relations particulières traduisant la composition d'un objet (la salle à manger de notre exemple est composée de la table et des chaises).

Par ailleurs, il convient de pouvoir gérer les ambiguïtés d'interprétation dues soit à l'imprécision du geste, soit à l'existence de relations spatiales ou fonctionnelles entre les objets candidats à la sélection, ou encore à l'apparition de difficultés à départager les candidats. Une solution possible est la mise en place de mécanismes permettant de construire un ensemble d'hypothèses d'objets candidats suffisamment large pour lesquels on définit une priorité traduisant qu'un objet a plus de chance d'avoir été désigné par le geste qu'un autre. Cet ensemble d'objets sera ensuite filtré par les informations langagières et les objets en surplus seront éliminés.

3.3.2. Interprétation dans le contexte du dialogue

A ce stade, nous pouvons mentionner le rôle que peut avoir le contexte de dialogue proprement dit dans le processus d'interprétation. Jusqu'à présent, nous avons fait l'hypothèse que l'analyse de la trajectoire gestuelle s'effectuait relativement à l'ensemble de l'espace graphique présenté à l'utilisateur. Pourtant, il apparaît que dans de nombreux cas il est nécessaire de considérer des sous-espaces de référence particuliers dont les éléments qu'ils contiennent seront seuls l'objet de l'opération de discrimination

associée à l'analyse d'une expression référentielle associée à un geste. De tels sous-espaces vont en général correspondre à des zones d'activité stables à une étape donnée du dialogue et qui peuvent être détectées à l'aide de différents indices : expressions référentielles utilisées dans les énoncés précédents, marques explicites d'un nouvel espace (par exemple : *Maintenant, on aménage le salon*), marques de continuité, structure de la tâche en général (découpage d'un appartement en pièces, regroupement perceptuel d'objets etc.). Dans l'exemple que nous avons choisi pour illustrer cet article, nous supposerons qu'il a déjà été tenu compte de tels phénomènes de focalisation, notamment pour décrire la structure des objets fournie à l'interprétation.

3.3.3. Stratégie d'interprétation

Comme nous l'avons vu, l'interprétation du geste en contexte permet d'effectuer le lien entre ce qui est visuellement perçu (la scène et le geste) et ce qui est perçu auditivement du point de vue d'un interlocuteur humain. Il semble intéressant de savoir dans le cas d'un traitement automatique dans quel ordre s'effectue cette prise en compte et dans quelles situations. De façon simplifiée, il est possible de dégager deux grandes stratégies correspondant à des configurations particulières de l'espace de présentation graphique :

a. une stratégie privilégiant le geste. Typiquement, il s'agira d'une tâche, ou d'une étape particulière au cours d'une tâche, pour laquelle les objets sont globalement de type homogène et donc pour laquelle le filtrage par la langue n'est pas déterminant. C'est par exemple ce que l'on rencontre dans ICPDraw où n'apparaissent que des figures géométriques [Caelen et al, 91]. Il est à noter que de telles tâches relativement simples facilitent l'emploi du déictique "décatégorisé" ça, opposition à des expressions démonstratives plus complexes.

b. une stratégie privilégiant l'expression référentielle. Celle-ci pourra être déclenchée lorsque l'espace de visualisation devient complexe de par le nombre important d'objets présentés, en association avec une grande variété de catégories. L'expression référentielle, par le filtrage qu'elle permet, va alors conduire à une simplification conceptuelle de la représentation pour ne sélectionner que les objets pertinents vis à vis du geste.

3.3.4. Retour à l'exemple

Pour chaque scène visualisée dans l'application, il est nécessaire de construire une représentation de ce que perçoit visuellement l'utilisateur. Il semble judicieux de rendre compte de la répartition des objets dans la scène, c'est ce qui est traduit par la figure 8 ou l'on tient compte de la superposition des objets. Ainsi, la scène peut s'appréhender de la manière suivante :

- la scène se compose de deux groupes d'objets distincts le salon et le canapé situé en dehors des limites de la pièce ;
- le salon comprend d'une part une salle à manger, d'autre part un tapis sur lequel se trouvent deux fauteuils et un poste de télévision.

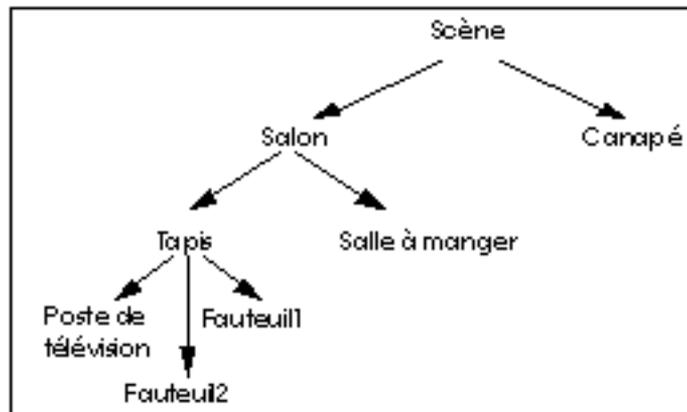


Figure 8 : Représentation de la répartition spatiale des objets dans la scène

Une stratégie d'interprétation possible pour le geste consisterait à repérer le nombre de syntagmes nominaux à valeur déictique pour éliminer les hypothèses de sémantique du geste non adéquates, puis ensuite à interpréter le geste en fonction de la scène.

Dans notre exemple, le message verbal *Déplace ces deux fauteuils ici* contenant un groupe nominal démonstratif et un déictique permet de conclure que le geste associé doit comporter au plus deux désignations. Ceci permet d'éliminer l'hypothèse 1 qui en contient 3 (voir figure 9.a).

La mise en relation de l'hypothèse de geste numéro 2 avec l'univers des objets conduit au résultat suivant (voir figure 9b) :

- la 1ère désignation permet de sélectionner les objets : {fauteuil1, fauteuil2, tapis} ;
- la 2ème désignation ne correspond à aucun objet, mais permet de sélectionner le lieu.

L'étude du groupe nominal démonstratif indique que les objets à prendre en compte, pour la première désignation, sont au nombre de deux et que ce sont des fauteuils. Ce qui correspond à la sélection des deux objets de type "fauteuil". Le déictique "ici" fait référence à un lieu, ce qui correspond à la sélection de la seconde désignation. Les objets étant complètement identifiés, la requête de l'utilisateur peut être exécutée.

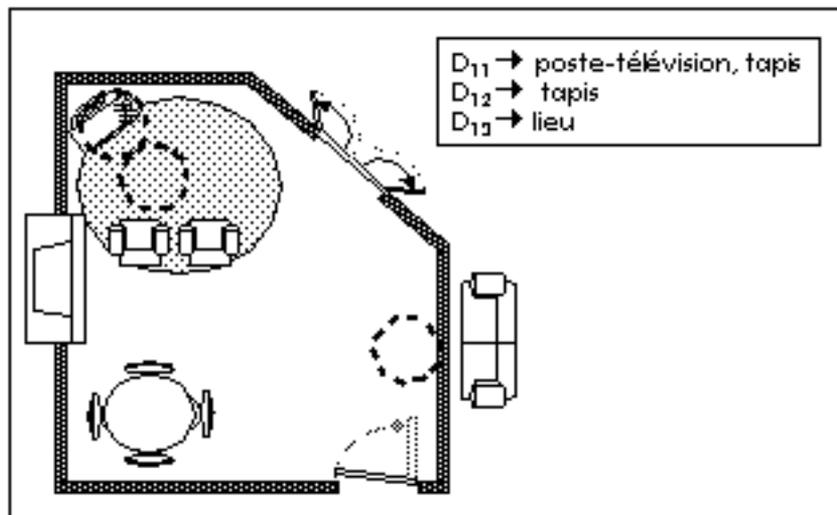


Figure 9a : Objets contrastés par le geste selon l'hypothèse 1

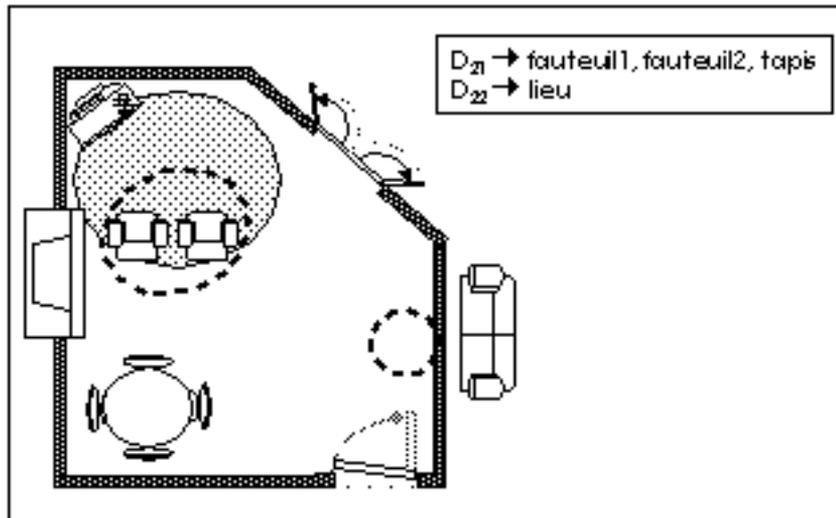


Figure 9b : Objets contrastés par le geste selon l'hypothèse 2

4. Conclusion

L'analyse que nous avons présentée ici n'est qu'une étape vers un modèle plus complet du geste dans une perspective de communication homme-machine. Elle s'inscrit dans le cadre de travaux plus généraux sur la référence, ce qui explique l'hypothèse faite *a priori* d'un geste dont la visée est essentiellement sémiotique en accompagnement de groupes nominaux énoncés par l'utilisateur. Par ailleurs, bien que les différentes étapes proposées pour l'analyse d'un geste aient conduit à des implémentations destinées à valider nos hypothèses, nous restons attachés à l'idée qu'il n'est pas sain que le développement technologique prenne le pas sur une certaine distanciation qui elle seule permettra de dégager de véritables modèles de la communication multimodale. Inversement, tout comme il apparaît impossible d'aborder les phénomènes de langue sans les lier à un moment ou à un autre aux circonstances d'élocution, c'est à dire là où l'interprétation va véritablement pouvoir s'opérer, il est essentiel de situer toute analyse du geste de désignation dans une tâche particulière, ou au moins dans une classe générique de tâche, de sorte à mettre en évidence des ensembles cohérents de phénomènes. A ce titre, les tâches de conception, qui sont étroitement liées à un domaine spécifique, sont de tout premier ordre pour servir de base à la mise en œuvre de véritables

systèmes de dialogue où langue et geste apportent plus de souplesse à la communication entre la machine et l'utilisateur, ce dernier pouvant ainsi consacrer toute son attention à l'activité de conception elle-même. Dans ce cadre, il reste malgré tout encore beaucoup de chemin à parcourir pour intégrer tous les paramètres qui interviennent dans l'interprétation d'un énoncé, qu'il s'agisse de caractéristiques perceptives, fonctionnelles et bien sûr linguistiques.

Bibliographie

- [**Bellalem, Romary, 93**] *Le dialogue homme-machine multimodal : vers la compréhension du geste de désignation.* Bellalem N. & Romary L., in *Actes de L'interface des mondes réels et virtuels, Montpellier, 24-26 mars 1993.*
- [**Cadoz, 92**] *Le geste canal de communication homme/machine la communication instrumentale.* Cadoz C., *Technique et sciences informatiques, Volume 13 - n°1, pages 31-61*
- [**Caelen et al., 91**] *Interaction multimodale autour de l'application ICPDraw.* Caelen J., Garcin P., Wret J., Reynier E., *Actes du Workshop IHM'91, Dourdan, Décembre 1991.*
- [**Corblin, 87**] *Indéfini, défini et démonstratif,* Corblin F., Droz, Genève-Paris.
- [**Cosnier et al., 82**] *Communications et langages gestuels.* Cosnier J., Berrendoner A., Orecchioni C., *Les voies du langage, Communications verbales gestuelles et animales, Edition Dunod-Bordas.*
- [**Gaiffe et al., 94**] *Références et gestion du dialogue.* Gaiffe B., Reboul A., Romary L., *Actes de TALN'94, Marseille.*
- [**Kleiber, 92**] *Y a-t-il un il ostensif ?.* Kleiber G. *Actas do XIX Congreso Internacional de Lingüística e Filoloxa Romnicas, Universidade de Santiago de Compostela, 1989, Publicadas por Ramn Lorenzo, A Corua, 1992.*
- [**Mignot et al.,93**] *An experimental study of future "natural" multimodal human-computer interaction.* Mignot C., Valot C., Carbonell N., *INTERCHI'93, Amsterdam.*
- [**Neilson, Lee, 94**] *Conversations with graphics: implications for the design of natural language/graphics interfaces.* Neilson I., Lee J. *Int. Journal of Human-Computer Studies, 40, 1994, pp. 509-541.*

- [Pribbenow, 93]** *Computing the meaning of localization expressions involving prepositions: the role of concepts and spatial context.* Pribbenow S., Mouton de Gruyter, Cornelia Zelinsky-Wibbelt (Editor), pages 441-470.
- [Romary, 93]** *«Mets ça ici» où quand «ici» dépend de «ça» : L'interprétation de «ici» dans des énoncés de positionnement.* Romary L., Workshop *Le dialogue homme-robot en langage naturel*, Caen, Octobre 1993.
- [Romary, 94]** *Sens et action, ou comment aménager son salon.* Romary L., in *Actes de TALN'94*, Marseille, 7-8 avril 1994.
- [Schang, Romary, 94]** *Frames, a unified model for the representation of reference and space in a Man-Machine Dialogue.* Schang D., Romary L., 3rd *International Conference on Spoken Language Processing*, Yokohama .
- [Wahlster, 91]** *User and discourse models for multimodal communication.* Wahlster W., *Intelligent user interface*, ACM Press Services, Sullivan J. and Tymler S. (Eds), Addison-Wesley .