Interface multimodale et conception

Jean Caelen
ICP-INPG, 46 av. Félix Viallet, 38031 Grenoble Cedex 1

RESUME. Les interfaces homme-machine multimodales peuvent apporter des améliorations aux systèmes de conception assistée par ordinateur (CAO) : (a) par une bonne adéquation des modes de communication à la tâche, (b) par le rendu des sorties et les possibilités du modelage gestuel, et plus généralement (c) par l'interaction de type dialogue en langage naturel avec les objets de la scène. L'environnement de CAO que nous proposons tente de rendre transparentes à l'utilisateur, les phases de conception et de modelage à travers un seul composant de dialogue. On vise ainsi à augmenter la richesse d'expression tant langagière que non-langagière et à résorber la charge de planification de l'utilisateur à travers le jeu du dialogue qui offre des facilités de communication et d'expression plus grandes. Nous posons ces problèmes, nous en donnons quelques solutions et nous illustrons notre propos autour de la plate-forme ICPPlan, logiciel multimodal pour la conception de plans architecturaux.

MOTS-CLES. Interface homme-machine, dialogue.

1. Introduction

La tâche de conception assistée par un ordinateur (CAO) peut être considérée comme une activité coopérative — supportée par le dialogue — entre deux partenaires : l'homme et la machine. Ces deux partenaires concourent à réaliser un objectif qu'ils construisent en commun en coordonnant leurs actions. Dans ces conditions, la machine doit non seulement collaborer à la tâche de conception proprement dite mais également se charger de la tâche de création et de composition graphique. Nous faisons l'hypothèse que la conception est une tâche de type innovante dans laquelle les stratégies de planification sont peu prédictibles ; c'est pourquoi le dialogue, ayant ses propres structures normées par l'usage, peut apporter un cadre structurant aux activités de conception (on conçoit souvent à plusieurs en utilisant le dialogue comme moyen de création collectif).

Généralement, dans les systèmes anciens, les phases de l'activité, conception et modélisation, bien que consécutives et bouclées l'une sur l'autre, étaient

fonctionnellement séparées ; le concepteur définissait son objet avant de le composer puis revenait y faire des retouches qui étaient considérées comme de nouvelles conceptions. Pour cela, les modules informatiques associés à ces deux fonctions s'échangeaient des informations complexes décrites dans un langage de description formel.

Notre approche est plutôt d'intégrer à travers un module de dialogue ces deux phases du travail en variant les présentations de l'objet et en augmentant les capacités d'interaction du concepteur sur l'objet composé. Nous nous démarquons par là des approches en CAO fondées sur les méta-modèles dont l'objectif est d'offrir au concepteur un processus dynamique permettant à la fois de représenter des connaisances sur les objets mais ausi sur le processus de conception (Kiriyama, 1989 : 429-449). Certes, ce méta-modèle permet de varier les différentes étapes du processus de conception mais nécessite un ordonnancement explicite de la part du concepteur ce qui entrave la création spontanée. Dans notre approche cet ordonnancement est implicite puisqu'il se construit au fur et à mesure de l'évolution du dialogue.

Notre objectif est également de concevoir une architecture d'interface homme-machine multimodale (utilisant les modes parole, geste, vision) donnant ainsi un pouvoir accru d'expression et de manipulation à l'utilisateur dans le cadre d'applications de CAO. La multimodalité, sur laquelle nous reviendrons plus loin, offre en effet des possibilités de présentations multiples tandis que les couches de communication et de dialogue implémentées dans l'interface permettent d'enrichir les capacités d'interaction et d'unifier les activités de conception et de composition (dans la suite de l'article, nous appelons cette dernière "modelage").

2. Environnements informatiques

Les interfaces multimodales à composante langagière sont possibles à l'heure actuelle car : (a) la puissance des machines et les environnements informatiques ont beaucoup évolué depuis une décennie, (b) les périphériques d'interaction — entrée vocale, tablette de griffonnage, gant numérique, etc. — ont atteint un degré d'efficacité et de précision tel que l'on peut désormais les envisager comme médias d'interaction.

2.1. De la modélisation géométrique à la modélisation déclarative

Les systèmes de CAO actuels permettent essentiellement de décrire des projets architecturaux dont on a une idée précise *a priori* (CATIA, EUCLID, AUTOCAD, etc.). Vers la fin des années 80, une génération de logiciels, à base de systèmes experts, a vu le jour soit pour reproduire des règles de construction, soit pour assister l'architecte dans la réalisation de plans. En fait, la modélisation de scènes tridimensionnelles par ordinateur est un processus de création (et de composition) visant à utiliser l'ordinateur comme un outil d'expression.

C'est pourquoi depuis le début des années 90, l'intérêt d'une approche déclarative pour la création et l'animation de scènes tridimensionnelles, par rapport aux

approches classiques s'est porté sur la description de la scène en termes de propriétés et de contraintes — à partir desquelles une représentation de la scène est calculée. Elle permet à un créateur ou à un architecte de décrire son projet. De nombreuses études ont vu le jour dans le but de donner à un utilisateur des outils qui lui permettraient de *représenter* une scène. On peut citer le projet ExploFormes (Lucas, 1989) dont l'objectif a été l'exploration de formes définies à l'aide d'une description déclarative. On mentionnera également :

• le projet MultiFormes de création de scènes (Plemenos, 1991),

• les travaux de (Djedi, 1991), (Gaildrat, 1993 : 265-284) qui ont utilisé une méthode déclarative pour la modélisation en synthèse d'images,

• et ceux de (Donikian, 1993) qui se distinguent des précédents en proposant une méthode de conception déclarative basée sur la création d'un scénario décrivant l'environnement du point de vue de l'usager. La scène est décrite, de ce point de vue, en termes de propriétés, de contraintes et de relations portant sur les objets.

Le lecteur pourra par exemple consulter (Balet, 1993) et (Donikian, 1993) pour une synthèse de ces différentes méthodes.

2.2. Vers une interface mieux adaptée : la multimodalité

2.2.1. Principes généraux

Ces dernières années ont connu une certaine invasion des interfaces grahiques à *manipulation directe* (Coutaz, 1990). Elles ne satisfont pas pleinement les contraintes de communication nécessaires à la conception (Buxton, 1993 : 21-22). Par exemple, il n'est pas possible de désigner des objets cachés et encore moins des objets qui n'existent pas encore dans la scène en cours de construction ; il n'est pas possible de différer, réitérer des commandes, etc. toutes choses pour lesquelles le langage naturel est bien adapté. Or ce sont là typiquement des nécessités dans une tâche de conception qui par essence, manipule une scène évolutive et complexe. Une composante de dialogue gérant des activités langagières orales et/ou écrites y est donc indispensable puisque seule la langue offre une puissance d'expression adaptée (Pierrel, 1989 : 91-112). Le langage permet effectivement d'énoncer des requêtes complexes d'un niveau d'abstraction élevé (Arnold, 1993) et de décrire des scènes. Les travaux de (Denis, 1993 : 497-506) qui abordent l'analyse des relations entre la cognition visuo-spatiale et le langage dans une tâche de description, montrent que le langage est par essence, de nature linéaire et unidirectionnelle et renccontre des difficultés à décrire des objets 2D. Cela pose donc la nécessité d'intégrer plusieurs modes de communication — langagiers et non-langagiers — et d'équilibrer leurs usages dans une tâche de conception. Ce sont là les objectifs et les problèmes d'une *interaction multimodale.*

Par définition, l'interaction multimodale (Coutaz, 1991 : 13-16) met en jeu simultanément et de manière coopérante, plusieurs canaux sensori-moteurs de l'être humain — vision, parole (entendue et produite), geste (mouvement, désignation, écriture, dessin), etc.. L'utilisateur est considéré alors en interaction plus "naturelle" avec la machine ; par exemple on doit pouvoir énoncer «*mets ça là*» en désignant par le geste un objet (->ça) et un lieu (->là). Cet exemple montre qu'il y a pour la

machine un problème de résolution de référence spatiale et temporelle entre les modes — oral et gestuel — et que l'interprétation des commandes multimodales est un processus complexe.

Ce type d'interaction nécessite donc la conception de nouvelles architectures logicielles (Collectif IHM'92, 1992), ayant deux fonctionnalités essentielles : (a) la gestion et l'interprétation des informations multimodales, (b) le contrôle du dialogue. En outre, la mise en œuvre de telles architectures logicielles nécessite la modélisation des éléments suivants :

- la connaissance de l'usager (niveau d'expertise, caractéristiques personnelles, etc.),
- la connaissance du domaine de la tâche (règles pour la conception et pour la création),
- le processus d'inférence des intentions et des buts de l'usager concernant le problème à résoudre,
- les règles de l'intervention pédagogique (aides, guides, exemples),
- les règles du dialogue et de la communication fondées sur des principes de négociation et de coopération.

2.2.2. Un modèle cible

En CAO, malgré la convivialité et la puissance des algorithmes de visualisation de rendus, certaines étapes demeurent encore fastidieuses. Par exemple pour créer la scène *Un verre sur la table* l'utilisateur devra raisonner en termes géométriques pour faire coïncider le fond du verre et le dessus de la table sur le même plan. Il serait en effet plus commode d'énoncer l'ordre *Pose le verre ici* en désignant la table sans se préoccuper de la génération des actions-machine correspondantes. C'est pourquoi il est utile d'étudier un système déclaratif doté d'une interface multimodale de scènes qui a pour objectif d'aider les architectes dans leur tâche de conception. Ce système déclaratif doit être pilotable par une interface permettant l'usage de différents modes d'entrée pour notamment, formuler des commandes de haut niveau d'abstraction en langage naturel, sans manipulation de données numériques comme par exemple le coefficient de luminosité, les coordonnées des objets, etc. Ce n'est pas le cas aujourd'hui car la création de scènes complexes reste encore un processus long et difficile qui impose au concepteur des opérations de bas niveau l'obligeant à un effort d'adaptation au logiciel.

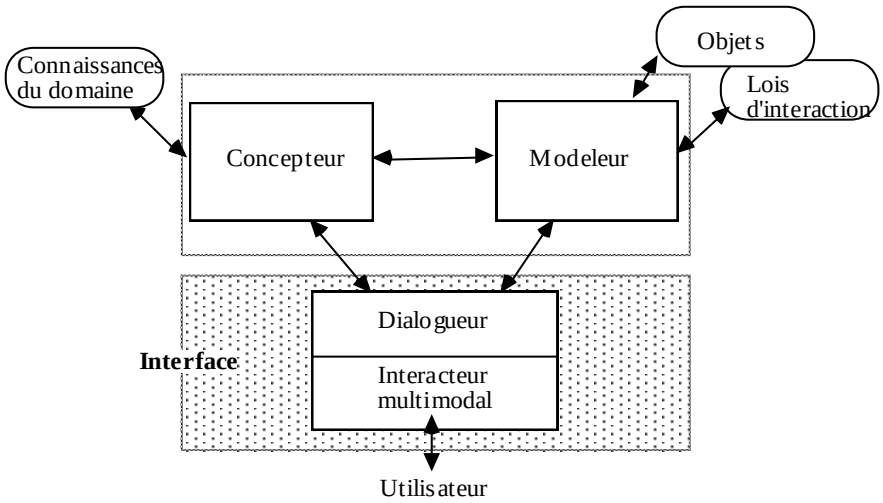


Fig. 1 : Synoptique du système de CAO doté d'une interface multimodale.

Pour pallier ces inconvénients nous avons tenté d'appliquer les nouveaux concepts de la multimodalité et les dernières avancées du dialogue homme-machine (Ozkan, 1993a : 77-84) à la CAO. Cette idée nous a conduit à l'organisation du système de CAO représenté à la figure 1. On peut y distinguer deux parties :

• l'interface composée de l'interacteur multimodal — qui collecte les événements multimodaux et les associe au contexte général de l'action — et du dialogueur, tous deux considérés comme des agents autonomes qui interprétent (ou générent en sortie) les énoncés multimodaux,

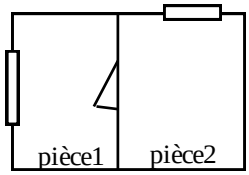
• le système de CAO constitué de deux sous-systèmes : le concepteur et le modeleur qui échangent, dans le sens concepteur-modeleur, des spécifications pour la présentation des objets et inversement, dans l'autre sens, des informations sur les objets traités.

Il faut assurer une certaine réactivité entre ces deux sous-systèmes hors du contrôle du dialogueur et les concevoir comme deux ensembles d'agents autonomes. De ce point de vue, le sous-système concepteur se limite au raisonnement sur les objets du domaine, ce qui implique que le système possède des connaissances architecturales, tandis que le modeleur assure toutes les fonctions de représentation spatio-temporelle des objets. Le rôle de l'interface est de mettre la machine en relation avec l'utilisateur, de structurer le dialogue et de distribuer les messages aux deux agents.

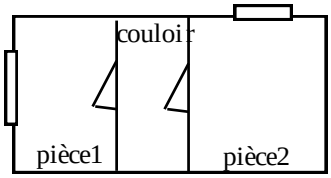
Pour illustrer l'intérêt de cette architecture, examinons à travers l'exemple ci-dessous la suite des échanges entre le modeleur (noté M) et le concepteur (noté C). Ils ont tous deux un comportement autonome c'est-à-dire qu'ils sont capables de travailler dans leur propre domaine de compétence sans recours extérieur. Ce n'est que lorsqu'il leur manque une information ou que la tâche qui leur est demandée dépasse leur compétence qu'ils font appel à des ressources externes. En adoptant une telle stratégie opportuniste l'ensemble du système peut mieux suivre les nombreux

changements de focus inhérents aux incidences et ruptures de plans si fréquents en conception.

Soit un scénario dans lequel l'utilisateur énonce : «*mets un couloir entre ces deux pièces*» et supposons qu'il ait en tête de conserver la surface des pièces. Supposons par ailleurs que le dialogueur (D) soit capable de "comprendre" puis d'interpréter un tel ordre tout en inférant correctement l'intention de l'utilisateur, le système doit passer de l'état (1) à l'état (2).



état (1)



état (2)

Ceci implique pour le système la suite des opérations suivantes :

0. C décompose l'ordre *ajoute(couloir, scène) ET scène(pièce1, pièce2)* que lui a adressé D, en actions élémentaires planifiables que M ou lui-même sait exécuter,

1. *C->M:* C envoie un message de haut niveau d'abstraction à M lui spécifiant une action élémentaire à exécuter, ici l'action *dilater(pièce2)*,

2. M exécute la commande et en cas de succès lui retourne un acquittement. Supposons que ce soit le cas, *M->C : acquittement*,

3. *C->M: dessine(cloison-séparatrice, pièce2)*,

4. M exécute la commande qui consiste à ajouter une cloison (trait d'une épaisseur connue pour M, mais dont la forme même échappe à C) entre les murs extérieurs de la pièce2. Il est de la compétence de M de régler le problème de leur intersection. Puis *M->C idendifieur(objet = couloir) ET attributs(objet)* qui donne à C toutes les informations architecturales utiles dont il a besoin concernant ce qui vient d'être fait par M,

5. C continue son raisonnement de nature experte. Par exemple *C->M : extension(couloir)* s'il juge que le couloir n'a pas une largeur suffisante,

6. M exécute la commande et en cas de succès envoie un acquittement, *M->C : acquittement*,

7. à ce stade on pourrait considérer l'opération terminée et rendre le contrôle au dialogueur. En fait une action vraiment pertinente serait de rajouter une porte sur la cloison, problème qui devrait être soulevé par C ou par M (pour des raisons architecturales ou de symétrie géométrique, etc.). Cela exige que le système prenne maintenant l'initiative en proposant une relance et en demandant confirmation à l'utilisateur par une incidence dans le dialogue. D'où l'échange *C (ou M)->D : milieu?(porte, cloison)*;

8. que D traduit en un message compréhensible par l'utilisateur via l'interacteur multimodal. Cela donnerait par exemple «*Faut-il dessiner une porte ici ?*» (le mot ici étant accompagné d'une zone clignotante sur l'écran).

Cet exemple illustre également le type de raisonnement synchronisé que doivent effectuer les deux agents C et M pour prendre en compte un énoncé préalablement analysé par le dialogueur. Ces mécanismes font de l'interface une *interface à base de*

connaissance multi-agents qui intègre des techniques issues de l'intelligence artificielle, de la CAO et des systèmes de gestion à base d'objets, sans oublier l'ingéniérie des interfaces et le dialogue homme-machine.

3. Etude expérimentale

Pour aboutir à une interface performante et utilisable, le concepteur d'interface (ou l'ergonome) fait généralement une analyse des besoins en partant de situations réelles. Nous avons adopté cette démarche d'autant plus que les tâches de conception sont cognitivement mal connues. Nous avons fait une étude des actes de description et de désignation vocales de figures spatiales abstraites et concrètes (éléments architecturaux, mobilier, etc.). Pour cela nous avons procédé à une expérience de type Magicien d'Oz permettant de simuler une interaction homme-machine en substituant la machine par un compère humain. Dans cette expérience un instructeur donnait des ordres à un manipulateur pour lui faire dessiner des figures de complexité sémantique croissante. Nous appelons tâche, la composition d'une scène.

3.1. Protocole expérimental

L'expérience présentée ci-après restreint le cadre de l'étude de la conception de deux manières : (a) d'abord, le domaine de la tâche est réduit à la description de figures spatiales simples ; (b) ensuite, les stratégies communicationnelles ne sont abordées que pour l'acte de désignation.

Le choix de cette focalisation est motivé par trois facteurs : (a) tout d'abord, la description d'une figure suit un plan décomposable en actes de désignation des éléments de cette figure; (b) en deuxième lieu la désignation est le point de convergence entre les modes verbal, gestuel et visuel, modes qui nous intéressent particulièrement pour la communication homme-machine multimodale ; (c) en troisième lieu il nous apparaît que le domaine de la désignation spatiale dans un cadre actionnel est particulièrement intéressant pour l'étude du langage opératif de conception.

La situation expérimentale était la suivante (Ozkan, 1993b : 99-108) : une série de figures était présentée à l'instructeur, qui donnait verbalement des instructions au manipulateur afin que ce dernier les exécute à l'ordinateur au moyen des périphériques classiques, la souris et le clavier. Le manipulateur ne connaissait pas les figures. Les productions verbales des deux agents étaient enregistrées pour analyse. Les figures données au sujet étaient réparties en trois groupes (fig. 2) : un premier groupe constitué de figures abstraites proposées par Levelt (Levelt, 1982a : 199-220), (Levelt, 1982b : 251-268) (fig. 2a), un second groupe contenant des figures structurellement identiques à celles du premier groupe, mais représentant des pièces reliées par des portes (l'icône en forme de losange - fig. 2b.), et un troisième groupe constitué toujours des mêmes figures, mais incluant également des icônes figurant des meubles et fenêtres (fig. 2c). L'expérience met donc en jeu trois mondes de complexité sémantique croissante et ceci de façon successive. Cela nous permet de partir des conclusions de Levelt pour l'interprétation des résultats de

l'expérimentation décrite ci-dessus.

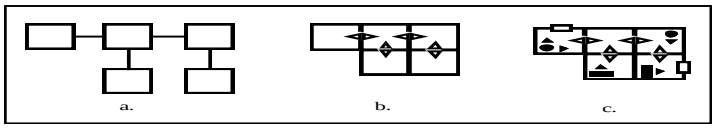


Fig. 2 : Les trois types de mondes : un instructeur fait dessiner des figures sur ordinateur à un manipulateur.

3.2. Quelques observations recueillies

Voici quelques exemples de productions verbales enregistrées au cours du dialogue et transcrites orthographiquement (en tenant compte des particularités de l'oral selon la notation utilisée par (Fréchet, 1992) pour indiquer les hésitations, les reprises, les souffles, les allongements de syllabes, etc.) :

E1 : «*On part du milieu on fait un carré,, voila oui voil non en bas en bas on part toujours d'en bas pour l'instant oui c'est vrai j'ai pas dit (h) et puis on trace une ligne toujours de la même manière au milieu du côté horizontal haut, (+) qui permet de placer le carré 1 toujours de la même manière aussi, voila voila une ligne° horizontale à droite, côté: droit oui, perpendiculaire toujours qui permet de placer le carré 7,, et en repartant, du carré, du milieu, voila une ligne horizontale à gauche qui va permettre de placer le carré*»

Dans cet exemple, l'énonciation du sujet est très imbriquée dans l'action du manipulateur; elle l'accompagne dans ses succès ou échecs.

Dans l'exemple suivant, le langage est rendu très opératif, récurrent même (séquence répétée "prends"+"pose") :

E2 : «*Prends un carré et pose-le dans le bas de l'écran (+) dessine un trait vertical et pose-le au milieu du haut du carré (+) prends un deuxième carré et pose-le au-dessus du trait (+) le trait vertical doit toucher le milieu du bord bas du carré (+) prends un trait horizontal et pose-le contre le milieu du bord gauche du carré 5 (+) prends un carré et pose le milieu de son bord droit contre le trait horizontal (+) prends un trait vertical et pose-le contre le milieu du haut du carré 9 (+) prends un carré pose le milieu du bas du carré contre le bord du trait vertical (+) prends un trait horizontal et pose-le contre le milieu droit du carré numéro 2 (+) prends un carré et pose le milieu de son bord gauche contre le trait horizontal* »

ce qui signifie que dès qu'un acte de parole a montré sa productivité vis-à-vis du manipulateur, le sujet l'utilise systématiquement, même si des incohérences par rapport au monde de référence sont attestées (les objets appartiennent à un monde géométrique artificiel, "carré", "trait", etc., tandis que les verbes "pose", "prends", etc., appartiennent à un monde métaphorique). L'essentiel pour le sujet semble être le seul succès de son acte de parole.

E3 : «*donc une pièce qui va être la pièce 8, qu'on place° au milieu en bas,, (+) dans cette pièce e on ** on rentre par une° porte simple sur du côté° sur le côté horizontal bas {M: donc j'place une porte-simple ?} <oui> au milieu du côté horizontal bas une fenêtre au milieu du° mur de gauche, {M: e pardon} (m) alors, une porte-double je ne pense pas qu'on puisse le dire donc il faut d'abord placer la*

pièce 5 (h) qui se trouve contigüe à la pièce 8 au-dessus alignée voila entre ces deux pièces une porte-double (h),, {M: ho} d'accord dans cette pièce 5 une fenêtre à droite au milieu du mur de droite... »

Dans ce dernier exemple, la complexité de la description verbale montre un enchevêtrement entre plusieurs niveaux de description : celui des objets référentiels, celui de l'organisation de l'activité et celui de la structure du discours. En effet, il est difficile au sujet de placer une "porte" entre deux "pièces" qui ne sont pas encore totalement dessinées. Il est, du fait de la linéarisation qu'impose le langage, également difficile au sujet d'exprimer clairement son plan. C'est donc ici que la multimodalité peut apporter une économie d'expression en offrant plusieurs canaux simultanés de communication.

3.3. Analyse des résultats de l'expérience

Une telle expérience est riche de plusieurs points de vue : représentation des connaissances, stratégies de dialogue et de communication et planification de la tâche.

3.3.1. Représentation des connaissances

Les tâches décrites ci-dessus peuvent être considérées comme étant de même nature mais plongées dans un monde sémantique différent, avec ses objets et ses actions spécifiques donnant lieu à des représentations mentales différentes chez le sujet. Bien qu'on retrouve des pièces dans les tâches 2 et 3, il est utile de les distinguer conceptuellement, parce qu'une pièce dans la tâche 2 n'a pas un rôle de contenant qu'elle n'acquiert que dans la tâche 3. Les connaissances à propos des objets varient donc selon la tâche.

Chez la majorité des sujets, les descriptions à la fin de la session sont beaucoup plus économiques que les descriptions en début de session. Certaines connaissances explicites deviennent partagées entre le sujet et le manipulateur au cours de leur interaction, pour s'assimiler au contexte commun d'interprétation des actes locutoires. Chez certains sujets, la progression des connaissances partagées sur les relations entre objets au cours de la session est très marquée. Comparons par exemple, les deux extraits suivants des productions verbales de la tâche 1 pour un même sujet. Le premier extrait se situe en début de tâche et le second en fin de tâche. Tous deux expriment la position d'un carré.

E4 : «*Tu choisis un troisième carré que tu viens placer dans l'alignement donc à la suite de ce segment horizontal et donc tu viens placer le côté gauche de telle façon que son milieu coïncide avec le segment de droite ...* »

E5 : «*Tu viens placer à la suite de ce segment de droite un nouveau carré donc le prolongement ...* »

Ces connaissances implicites en fin de tâche concernent la complétude de la relation définie entre les éléments de la figure. Seule la direction est jugée suffisante et nécessaire par le sujet pour définir la relation. Si nous nous penchons sur les attributs implicites, nous observons que :

• la relation peut être inférée à partir de la définition de classe des arguments (contenant ou contenu),

• les distances en x et en y entre les arguments sont toujours les mêmes d'une figure à l'autre,

• le référentiel est toujours le même au cours d'une tâche.

Ainsi, nous remarquons que les connaissances répétitives ainsi que les connaissances pouvant être inférées à partir d'éléments par ailleurs connus deviennent partagées par l'instructeur et le manipulateur.

3.3.2. Changement de monde de référence

Un énoncé peut renvoyer à plusieurs mondes de référence. Il s'agit là d'un phénomène communicationnel courant qui ne provoque pas de rupture d'interprétation si les mondes de référence sont compatibles. Dans le cas de notre expérience, les tâches se succédant dans le temps constituent des mondes distincts et exclusifs. Nous observons toutefois que pour l'énoncé E6 relatif à la tâche 2,

E6 : « *... on place une première porte au niveau du côté inférieur de ce* ***carré*** *au milieu voilà ...* »

le mot en italique n'appartient pas au monde de la tâche en cours, car, strictement parlant, l'objet ***Carré*** n'existe pas à la tâche 2. Ceci est attribuable ici, comme dans la majorité des cas observés, au phénomène d'amorçage, où un objet appartenant à un monde référencé lors d'une tâche précédente, reste présent dans la mémoire du sujet.

L'apparition, dans les productions verbales, d'un objet nouveau, qui n'appartient à aucun des mondes des tâches, constitue un autre type d'enchevêtrement de mondes de référence. Par exemple :

E7 : «*et pose-le au-dessus du carré numéro 3 avec un décalage d'un* ***carreau*** *...* »

Le ***carreau*** en question est un élément du quadrillage de fond de la zone de travail. Ces objets nouveaux peuvent appartenir à un monde quelconque faisant ou non partie du monde de la tâche. Ce sont ces phénomènes de changement de mondes de références qui sont à la source de ruptures dialogiques qu'il faut prendre en compte dans la modélisation du dialogue pour le domaine de la CAO.

3.3.3. Planification de la tâche

L'analyse de la séquence des énoncés permet de constituer le plan d'activité du sujet. La constitution du plan d'activité facilite l'analyse des stratégies cognitives de description. Dans ce domaine, rappelons certains des résultats qui nous intéressent particulièrement. Levelt (Levelt, 1982a : 199-220) a distingué deux grands groupes de stratégies de description : la description structurelle de la figure et la linéarisation de la figure par la description successive de chaque noeud selon les liens. En conception cependant, le plan d'activité indique que la stratégie du sujet n'appartient ni à l'une ni à l'autre de ces catégories, mais qu'elle est mixte. Il s'agit en fait d'une décomposition de la figure en trois parties, décrites successivement par linéarisation.

La constitution du plan d'activité facilite également le repérage de l'évolution des stratégies de linéarisation en cours de tâche, et donc de l'apprentissage du sujet. La stratégie du sujet en début de tâche est la linéarisation qui peut se qualifier de stratégie opportuniste du fait qu'elle ne requiert pas de planification. En fin de tâche, la description est basée sur les régularités observées de la figure par le sujet. De fait, la grande majorité des sujets passent d'une stratégie opportuniste à une stratégie de planification de la description en fonction des caractéristiques structurelles de chaque

figure.

4. Cahier des charges pour une interface multimodale dédiée à la conception

Il ne s'agissait pas dans le paragraphe précédent d'une véritable tâche de conception mais d'une sous-tâche liée à la désignation-description qui montre déjà la complexité des interactions et des comportements différents selon que les objets sont statiques, composites ou en mouvement. Il est évident que la richesse des communications entre humains fondée sur un implicite important et sur des connaissances d'arrière-plan ne pourra pas être reproduite en machine. L'analyse des expérimentations présentées dans la section précédente entraîne la prise en considération des aspects — mondes de références, modèle de dialogue, apport de la multimodalité et architecture — pour la conception d'interface homme-machine.

4.1. Mondes de référence distincts

Chez le concepteur architecte le problème de la rémanence et de l'amorçage entre les objets de mondes de référence différents se pose aussi puisqu'il a plusieurs sous-tâches distinctes à faire (l'esquisse, l'avant-projet sommaire, l'avant-projet détaillé et le projet final) et qu'il opère dans des domaines distincts (gros-œuvre, charpente, etc.).

Une interface ergonomique doit donc présenter les mondes de référence de façon claire et spécifique. Elle doit également minimiser les enchevêtrements entre ces mondes. Ceci nécessite de répertorier les mondes de référence qui sont pertinents pour les utilisateurs tels qu'ils se manifestent dans leurs énoncés. Une solution technique consiste à représenter les objets dans ces mondes à l'aide de réseaux sémantiques (Bourguet, 1992b : 369-374) ou de traits sémantiques dans les objets (Gaildrat, 1993 : 265-284), (Balet, 1993) selon les approches méthodologiques Cela permet par exemple en parcourant le lien *forme-de* qui relie les objets "pièce" et "carré" d'inférer les connaissances implicites comme dans l'énoncé E6.

4.2. Modèle de dialogue

La structure des échanges entre le manipulateur et le sujet nous amène à élaborer un modèle du dialogue homme-machine spécifique aux tâches de conception. Ce modèle est basé sur une analyse des récurrences des ruptures dans l'interaction et présente plusieurs stratégies de dialogue possibles (Ozkan, 1992a : 77-84). En effet, il est souhaitable d'éviter les ruptures dans l'interaction homme-machine — particulièrement en CAO — pour ne pas interrompre l'utilisateur dans sa réflexion et son geste créatif. En général, ces ruptures sont attribuables à des connaissances supposées par l'utilisateur mais non véritablement représentées dans la machine. L'analyse de l'évolution des connaissances partagées a fait ressortir deux des mécanismes par lesquels la connaissance devient implicitement partagée, donc potentiellement supposée. Nous en tirons deux moyens de retrouver les connaissances supposées afin d'éviter certaines ruptures en parcourant :

• les liens sémantiques entre des objets de différents mondes de référence,

• les historiques des connaissances récurrentes associées à l'interaction pendant la session en cours.

Nous aboutissons ainsi à l'idée d'un dialogue constructif (Ozkan, 1992a : 77-84), fondé sur l'enrichissement progressif et réciproque des connaissances des deux partenaires (homme et machine). D'un autre côté, la constitution automatique du plan d'activité à partir des énoncés permet également d'éviter certaines ruptures. La machine peut alors inférer les intentions de l'utilisateur à partir des plans d'activités successifs comme dans (Litman, 1985), (Nerzic, 1993).

4.3. Multimodalité

La multimodalité en entrée et en sortie est hautement souhaitable pour une tâche de conception afin de libérer l'utilisateur de toutes les entraves de la manipulation et de la désignation des objets qui le détourneraient de sa tâche créative proprement dite. La parole est un mode très compatible avec le geste de griffonnage (Faure, 1993 : 171-180) pour l'esquisse par exemple, mais aussi avec le geste de désignation. La multimodalité favorise également les activités multifils en rompant la linéarité du discours.

4.4. Architecture logicielle

La figure 1 présente une architecture logicielle qui répond au cahier des charges. Elle doit cependant se voir comme une représentation canonique. En réalité, l'importance relative des trois composants du système varie en fonction de l'étape de conception dans laquelle on se trouve et des objets qu'on manipule : cela va des grands ensembles de bâtiments aux constituants de base (matériaux) en passant par les espaces architecturaux (appartement, pièce, escalier, etc.) et les éléments d'architecture (mur, cloison, etc.). Pour certaines étapes, notamment le projet final, le composant de modelage doit être performant ; par contre pour l'esquisse le composant de dialogue doit être prédominant. C'est le cas d'ICPPlan qui accorde une place importante au dialogue — il s'adresse donc plutôt à l'étape de l'esquisse ou de l'avant-projet sommaire. ICPPlan peut être cconsidéré comme une instance de l'architecture logicielle générique.

5. ICPplan : conception par un dialogue multimodal

ICPPlan (Bourguet, 1992b : 369-374) est un logiciel qui offre les outils nécessaires à la réalisation d'esquisses de plans architecturaux au moyen d'un dialogue multimodal. Pour la réalisation d'ICPPlan les efforts ont surtout été mis sur le développement de l'interface (fig. 3) comprenant le dialogueur (D) et l'interacteur multimodal (IM), pour lesquels nous avons le plus de compétences.

Au regard du cahier des charges, l'agent modeleur a donc simplement les fonctions d'un éditeur graphique multimodal. Le module concepteur se compose d'un modèle de tâche dynamique et de quelques règles architecturales simples telles que :

< une pièce est entourée par 4 murs dont au moins l'un d'entre eux donne sur l'extérieur. Un mur extérieur peut recevoir une fenêtre. Il y a au moins une porte par pièce >. Ces connaissances sont représentées par un réseau sémantique, c'est-à-dire par un graphe dont les nœuds sont les objets du domaine (pièce, table, escalier, etc.) et les arcs, les relations entre les objets (partie-de, sur, à-coté-de, forme-de, etc.). Nous ne disposons donc pas d'un véritable système expert du domaine mais d'un système réduit assurant des fonctions similaires.

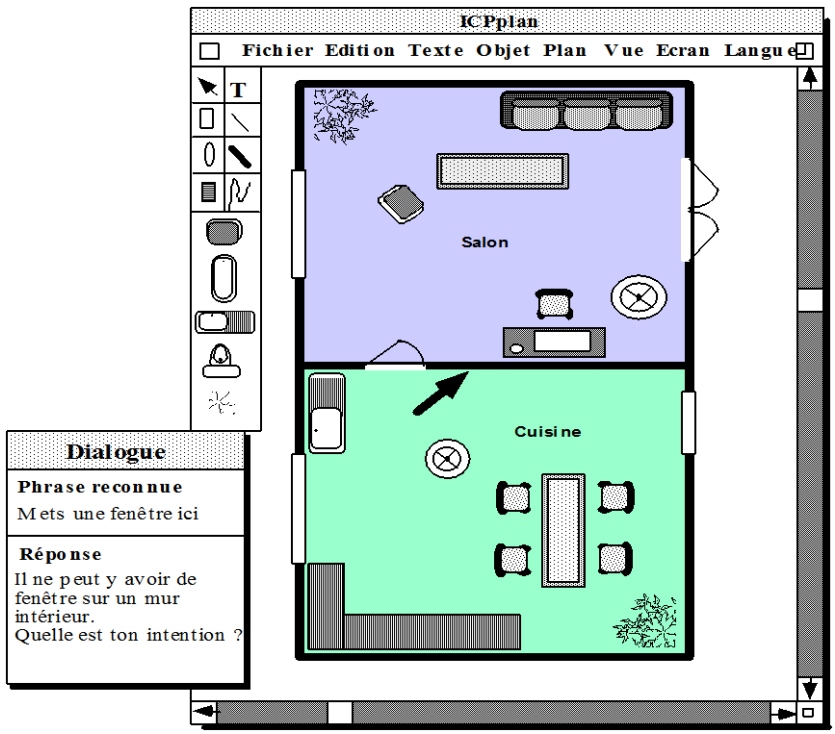


Fig. 3 : Vue d'une partie de l'interface représentant un plan 2D d'une maison. A cet instant l'utilisateur U tente d'effectuer une opération en désignant le mur de séparation de la cuisine par le curseur de la souris. Le dialogue se déroule de la manière suivante : U « mets une fenêtre ici » + geste, M « il ne peut y avoir de fenêtre sur un mur intérieur. Quel est ton intention ? ».

5.1. Puissance d'expression

L'interacteur multimodal offre à l'utilisateur les possibilités d'émettre des énoncés multimodaux tels que :

• «*Dessine une fenêtre. Agrandis la*» en langage oral ou écrit — ce type d'acte soulève des problèmes linguistiques de résolution de référence (anaphores). Ils sont traités par le dialogueur qui dispose d'analyseurs du langage naturel

• «*Mets la chaise à côté de la petite table*» qui est un problème de relation spatiale entre la chaise et la table,

• «*Place une porte sur ce mur*» qui combine désignation verbale et désignation gestuelle,

• les reprises, nombreuses à l'oral, comme «*dessine une petite table… non une grande*» sont également autorisées.

Le vocabulaire est actuellement de 300 mots ce qui semble suffire pour cette application. L'étude expérimentale a montré en effet que les utilisateurs focalisent leur lexique après la phase d'apprentissage du système. La puissance d'expression langagière n'est donc a priori limitée que par la puissance des analyseurs linguistiques. Pour l'interaction gestuelle nous nous sommes limités volontairement aux gestes de désignation au moyen de la souris (pointage précis, pointage vague, entourage).

En sortie, la synthèse de la parole et les messages écrits viennent compléter ou souligner les réponses graphiques de la machine. Les questions adressées à l'utilisateur comme «*je ne comprends pas*» sont émises simultanément dans les deux modes.

5.2. Fonctionnalités du système

a) du point de vue de l'interaction

Tous les énoncés émis par l'utilisateur sont interprétés par le dialogueur. Un retour visuel est prévu pour certaines actions (sélectionner un objet, le déplacer, etc.). Le dialogueur règle les problèmes d'incompréhension ou d'erreurs, ou traduit la commande ou la requête de l'utilisateur en un schéma d'actions possibles. Ce schéma est adressé au composant concepteur ou au composant modeleur en vue d'une exécution, conformément aux principes énoncés dans les paragraphes précédents. Le contrôleur dispose de règles et de méta-règles qui lui permettent de varier les stratégies de dialogue. Ces stratégies sont de type *directif, ou réactif, ou négocié, ou coopératif* (Bourguet 1992c).

b) du point de vue des capacités de raisonnement

L'agent concepteur effectue des raisonnements de nature géométrique (par exemple, les traits représentant des murs extérieurs ne peuvent se chevaucher) et des raisonnements de nature pragmatique. Ces derniers sont de type "parcours de réseau sémantique" et de type application de règles ou de contraintes. Par exemple, l'énoncé «*mets çà sur la table*» ou le «*çà*» réfère une porte est détecté comme incorrect parce qu'il n'y a pas de lien sémantique étiqueté "sur" entre "porte" et "table". Le système réagit par la réponse «*une porte doit être mise sur un mur*».

c) des fonctions de base de modélisation

Les fonctions du modeleur permettent la visualisation de plans 2D. Elles correspondent aux fonctions de base des éditeurs classiques de dessin augmentées de facilités pour nommer un élément en le désignant gestuellement, obtenir des aides sur la conception ou le dessin et l'historique des actions.

5.3. Limitations du système et perspectives

La stratégie de gestion des événements et de fusion des informations mise en œuvre dans l'interacteur multimodal, empêche ICPPlan d'être utilisé pour plusieurs fils d'activité (Bourguet, 1992a : 124-134). La solution proposée au problème de la coréférence entre les modes et au problème du parallélisme des médias (effets d'anticipation, de redondance, de conflit, etc.) limite quelque peu les possibilités d'interaction multiple avec les objets de la scène. Récemment ICPPlan s'est élargi au collecticiel (plusieurs personnes peuvent travailler en collaboration et simultanément sur l'application).

6. Conclusion

Cette première plate-forme ICPPlan ouvre la voie à de nouvelles approches en conception assistée par ordinateur. On relèvera parmi leurs caractéristiques :

- l'existence d'une grande synergie des deux agents conception-modelage,
- la puissance d'expression apportée par la multiplicité des modes et le composant dialogue.

L'état d'avancement de cette plate-forme rend envisageable l'évaluation de ses caractéristiques auprès d'utilisateurs. Pour notre propre compte l'objectif n'était pas de développer un environnement de CAO mais de démontrer pour ce domaine, la faisabilité d'une interface multimodale générique et d'en définir les spécifications d'une manière suffisamment précise.

Les principaux points qui restent à développer à l'heure actuelle sont :

- le composant concepteur qui devrait être un véritable expert du domaine et autour duquel d'amples collaborations sont à tisser,
- la complexification du modèle de dialogue pour tenir compte de l'entrelacement d'activités de différents niveaux inhérent aux tâches de conception et de modelage.

La poursuite des recherches sur le dialogue à plusieurs fils d'activités soulève à son tour des problèmes concernant :

- la gestion du focus, plus particulièrement le basculement d'un énoncé à un autre et la simultanéité dans la formulation d'énoncés indépendants,
- la gestion des informations partagées par les tâches qui se déroulent en parallèle dans chaque fil d'activité.

Les environnements graphiques tels que Windows 3.1. disposent de mécanismes de base qui offrent la possibilité de partager des connaissances entre les différentes activités. Cependant aucune boîte à outils ne rend compte dans un fil d'activité de l'effet de l'action réalisée dans un autre. Ces dialogues à multifils d'activité véhiculent des recommandations et des contraintes qu'il faudra considérer dans les modèles conceptuels des nouvelles architectures.

Remerciements

Nous remercions les participants à ce projet, à savoir, M.L. Bourguet, A.L. Fréchet et N. Ozkan qui ont permis par leurs réflexions, leurs discussions et les développements de logiciels, d'obtenir l'état actuel d'avancement de la plate-forme ICPPlan. Ce projet a été soutenu par le GDR-PRC "Communication Homme-Machine", le Conseil Régional Rhône-Alpes et le MESR.

7. Références

Arnold, M. (1993). L'ordinateur comme interlocuteur : problèmes posés par les interfaces utilisateur-machine, Rapport Interne, Lutèce'IA, Paris.

Balet, 0. (1993). Contribution à la conception et à la réalisation d'un modeleur déclaratif, Rapport de DEA IHS2M, Université Toulouse III.

Bourguet, M.L. & Caelen, J. (1992a). Interfaces Homme-Machine multimodales : gestion des événements et représentation des connaissances. Actes du colloque Ergonomie et Informatique Avancée, Biarritz, Octobre 1992.

Bourguet, M.L. (1992b). ICPPlan : dialogue multimodal pour la conception de plans architecturaux. Actes du Colloque des 19èmes Journées d'Etudes sur la Parole, Bruxelles, 19-22 mai 1992.

Bourguet, M.L. (1992c). Conception et réalisation d'une interface de dialogue personne-machine multimodale. Thèse de l'Institut de la Communication Parlée de Grenoble, option Signal Image Parole, Grenoble.

Brison, E. (1993). Interprétation des événements multimodaux. Rapport de DEA IHS2M, Université Toulouse III.

Buxton, B. (1993). HCI and the inadequacies of direct manipulation systems. SIGCHI Bulletin, 25 (1), 21-22.

Caelen, J. (1992). Compte-Rendu du "Workshop IHMM organisé par le GDR-PRC Communication Homme-Machine" à Dourdan les 13 & 14 avril 1992. Actes des 4ièmes Journées sur L'Ingénierie des Interfaces Homme-Machine, TELECOM Paris, 30 Novembre -2 Décembre 1992.

Collectif, IHM'92. (1992). Interfaces MultiModales et Architecture Logicielle, Rapport de l'Atelier Architecture Logicielle. (TELECOM Ed.), TELECOM Paris, 30 Novembre - 2 Décembre 1992.

Coutaz, J. (1990). Interface homme-ordinateur : conception et réalisation. Paris, Dunod Ed., 1990.

Coutaz, J. & Caelen, J. (1991). L'opération de recherche concertée homme-machine multimodale. Actes des 2ièmes Journées Nationales du GRECO-PRC EC2 Ed., Toulouse, 29-30 Janvier 1991.

Denis., M. & Cocude., M. (1993). Structural properties of visual images constructed from poorly or well-structured verbal descriptions. Memory and Cognition, 20, 497-506.

Donikian, S. (1993). Une approche déclarative pour la création de scènes tridimensionnelles: application à la conception architecturale, Thèse de docteur d'Université. Rennes I, 1993.

Faure, C. & Julia, L. (1993). Interaction homme-machine par la parole et le geste pour l'édition de documents : TAPAGE. Actes du Colloque L'interface des Mondes Réels & Virtuels, EC2 Ed., Montpellier, 22-26 Mars 1993.

Fréchet, A.L. (1992). Analyse linguistique d'un corpus de dialogue homme/machine Thèse de doctorat, Paris III.

Gaildrat, V. & Caubet, R. & Rubio, F. (1993a). Conception d'un modeleur déclaratif de scènes tridimensionnelles pour la synthèse d'images. Actes de la douzième conférence

Internationale sur la CFAO, l'Infographie et les Technologies Assistées par Ordinateur, MICAD'93 Hermès Ed., Paris, 9-12 Février 1993.

Levelt, W.J.M. (1982a). Linearization in describing spacial networks In S. Peters & E. Saarinens (Eds.), Processes, beliefs and questions. D. Reidel Publishing Company.

Levelt, W.J.M. (1982b). Cognitive Styles in the Use of Spacial Direction Terms. In R.J. Jarvella et W. Klein (Eds.), Speech, Place and Action. John Wiley and Sons.

Litman, D. (1985). Plan Recognition and Discourse Analysis : An Integrated Approach for Understanding Dialogue, Doctoral Dissertation and Technical Report 170. University of Rochester, NY.

Lucas, M. & Martin, D. & Martin, Ph. & Plemenos, D. (1989). Le projet ExploFormes, quelques pas vers la modélisation déclarative de formes. Actes du Colloque GROPLAN, Septembre 1989.

Nerzic, P. (1993). Erreurs et échecs dans le dialogue oral homme-machine. Thèse d'Université, Rennes I, 1993.

Ozkan, N., & Caelen, J. (1993a). Vers un modèle de dialogue adaptatif. Actes du Colloque L'interface des Mondes Réels & Virtuels (EC2 Ed.), Montpellier, 22-26 Mars 1993.

Ozkan, N., & Caelen, J. (1993b). La désignation spatiale : Une analyse pragmatique pour l'interaction homme-machine. Actes du Colloque Interdisciplinaire du Centre National de la Recherche Scientifique, Paris, 1-2 Avril 1993.

Pierrel, J.M., & Sabah, G. (1991). Dialogue en langage naturel écrit et oral : bilan des approches du CRIN et du LIMSI. Actes des 2ièmes Journées Nationales du GRECO-PRC, EC2 Ed., Toulouse, 29-30 Janvier 1991.

Plemenos, P. (1991). Contribution à l'étude et au développement des techniques de modélisation, génération et visualisation de scènes. Le projet MultiFormes, Thèse de doctorat d'état, Université de Nantes.

Salisbury, M.W. & Hendrickson, J.H. & Lammers, T.L. & Fu, C., & Moody, S.A. (1990). Talk and Draw : Bundling Speech and Graphics. IEEE Computer, 23 (8), August 1990.

Veerkamp, P. & Kiriyama, T. & Xue, D. & Tomiyama, T. (1990). Representation and implementation of design Knowledge for Intelligent CAD : Theoretical Aspects, Rapport N° CS-R9070, CWI.

Vigouroux, N. & Gaildrat, V. & Caubet, R. & Pérennou, G.(1992). Une architecture d'interface pour l'interprétation d'informations multimodales : application à un modeleur déclaratif de scènes. Actes des 4ièmes Journées sur L'Ingénierie des Interfaces Homme-Machine. TELECOM Paris, 30 novembre-2 décembre 1992.