

Un agent conversationnel pour naviguer dans les mondes virtuels

de Christophe Godéreaux, Pierre-Olivier El Guedj, Frédéric Revolta et Pierre Nugues, Caen (F)

Membres du laboratoire GREYC, commun à l'Institut des Sciences de la Matière et du Rayonnement et à l'Université de Caen, Caen (F)

1. Introduction

Nous présentons ici un agent conversationnel pour naviguer dans un monde virtuel. La navigation se révélant être l'un des points les plus délicats pour les nouveaux utilisateurs de ce type d'environnement. Notre prototype se constitue d'un dispositif de reconnaissance vocale du commerce ainsi que d'un circuit de synthèse de parole. Il s'appuie sur une architecture modulaire reliée à un environnement virtuel. Les entités du prototype ont pour rôle de traiter l'analyse syntaxique et sémantique, ainsi que le dialogue et les actions qui en résultent (Nugues, 1993; Nugues, 1994). Elles permettent à l'utilisateur de se déplacer à la voix dans un monde relativement complexe. Le système résultant s'intégrera dans un outil de téléconférence. Ce projet a été développé dans le cadre du programme COST-14 de la Commission des communautés européenne sur le travail coopératif à l'aide de l'ordinateur (COTECH, 1995). Dans ce COTECH, nous étudions plus particulièrement les outils linguistiques pour intégrer des possibilités de dialogue oral avec des assistants virtuels.

2. La réalité virtuelle et le travail coopératif

La réalité virtuelle est une des composantes des techniques multimédia et des autoroutes de l'information. Elle permet à un utilisateur la visualisation et l'interaction avec des données informatiques. Les systèmes de réalité virtuelle sont très dépendants du niveau technique de leurs différents composants. En effet, pour procurer des sensations réalistes, ces systèmes doivent simuler des scènes parfois complexes et interagir avec l'utilisateur en temps réel. La contribution des interfaces d'entrées-sorties dans leur amélioration est fondamentale. La qualité de la communication entre l'homme et les mondes virtuels dépend souvent de leur choix et de leur efficacité.

Les recherches sur le travail coopératif à l'aide de l'ordinateur – *Computer Supported Cooperative Work (CSCW)* – tentent de déterminer comment l'informatique peut aider des groupes de gens à travailler ensemble sur un projet, pour élaborer un produit, pour prendre une décision, etc. (Grudin, 1994). Le travail coopératif à l'aide de l'ordinateur fait parfois appel à des techniques de réalité virtuelle, par exemple pour créer des métaphores de salles de réunion. La société Xerox l'utilise pour structurer les interfaces graphiques (Robertson, 1993). Ces applications impliquent un accès multi-utilisateurs, à travers un réseau, à un même monde virtuel.

L'utilisation de métaphores de travail au bureau, comme des salles de réunion virtuelles dans certaines applications de téléprésence (Benford, 1993), peut aider à l'interaction dans un groupe en la rendant plus naturelle. Les techniques informatiques sont alors un intermédiaire pour des interactions entre personnes qui partagent un espace de travail virtuel.

Les métaphores utilisées ne permettent cependant pas, et probablement jamais, la même commodité et la même facilité que les interactions du monde réel. Cependant nous pensons qu'il est possible de les améliorer dans une large mesure et de faciliter l'accès à tous à ce type de travail. Pour rendre l'accès plus facile à ce type de mondes virtuels, l'interaction vocale ou linguistique est sans doute l'une des plus naturelle.

3. La réalité virtuelle et le dialogue oral

3.1. Pourquoi une interface de dialogue oral dans un monde virtuel

Des interfaces orales commencent à apparaître dans les systèmes de simulation ou dans les mondes virtuels (Allen, 1994; Ball, 1995; Bolt 1980; Karlgren, 1995). Elles sont à ce jour moins répandues que d'autres interfaces et très peu d'applications sont parvenues à un stade commercial. Ceci est sans doute dû au fait que ces interfaces sont très dépendantes du bon fonctionnement de la reconnaissance vocale dont la fiabilité n'était pas satisfaisante jusqu'à il y a peu.

L'interaction orale est pourtant un enjeu et permet d'améliorer sur certains points le réalisme des environnements virtuels. D'une manière générale, trois notions sont incluses dans les mondes virtuels (Quéau, 1993) : l'immersion, la navigation et l'interaction. Une interface de dialogue oral dans ces mondes facilite et amplifie ces trois notions pour l'utilisateur (Godéreaux, 1994).

L'immersion fait référence aux techniques d'immersion physique dans l'image grâce aux casques stéréoscopiques. Une interface de dialogue oral amplifie cette notion en ajoutant à cette immersion dans l'image, une immersion dans le son et la parole.

La notion de navigation traduit la possibilité d'évoluer, de se rencontrer, dans des univers virtuels. La navigation dans ces mondes virtuels est par essence complexe puisqu'elle se situe dans un espace à trois dimensions. Les systèmes de réalité virtuelle disposent de périphériques d'entrée, telles que la souris, le clavier, les manches à balai..., permettant de se déplacer dans un environnement tridimensionnel. Elles demandent à l'utilisateur un certain temps d'adaptation avant de pouvoir naviguer aisément. Ces interfaces permettent des déplacements rapides, lents ou précis dans un plan donné. Les déplacements sont cependant plus difficiles lorsque l'on désire «s'aligner» par rapport à un objet du monde; par exemple, pour se placer devant un tableau ou une porte permettant d'accéder à un autre monde. Ceci est d'autant plus difficile que l'objet ne se trouve pas dans le plan principal de déplacement.

Le dialogue oral par commandes simples facilite la navigation. L'adaptation de l'utilisateur est immédiate dès que, par exemple, un dialogue présenté par la fig. 1, est possible. Un dialogue plus élaboré comme celui de la fig. 2 permet de «s'aligner» avec précision par rapport à un objet du monde.

Système	<i>Pour vous déplacer vous pouvez dire, avance, recule, monte, descend, gauche, droite, stop.</i>
Utilisateur	<i>avance</i>
Utilisateur	<i>monte</i>
Utilisateur	<i>stop</i>

fig. 1 Un dialogue simple.

Utilisateur	<i>Place-toi devant le tableau.</i>
Utilisateur	<i>Vers quel autre monde puis-je aller?</i>
Système	<i>Vers le monde conférence ou le monde nature.</i>
Utilisateur	<i>Place-toi devant la porte du monde conférence.</i>

fig. 2 Un dialogue avec alignement.

La notion d'interaction fait référence à la manipulation de l'image, à sa transformation. Un grand nombre de modeleurs 3D pour les systèmes de réalité virtuelle existe déjà. Une interface de dialogue oral peut faciliter l'utilisation de ces modeleurs dans la désignation des objets puis dans leur manipulation. De plus il est aisé pour l'utilisateur de combiner navigation et interaction en s'exprimant oralement.

D'une manière générale, une interface de dialogue oral dans les mondes virtuels est un complément des interfaces classiques même si elle est parfois moins précise. Notamment, elle facilite la tâche de l'utilisateur dans la découverte de ces univers.

3.2. *Un système de réalité virtuelle pour le travail en groupe : DIVE*

DIVE, *Distributed Interactive Virtual Environment*, est un environnement virtuel réparti. Il se fonde sur le système d'exploitation Unix et les protocoles Internet (Andersson, 1994). Il a été développé par le Swedish Institute of Computer Science (SICS), membre de COST 14. Il a notamment comme but d'être un outil de téléconférence. Chaque utilisateur de DIVE, connecté au réseau Internet, est représenté graphiquement par un agent virtuel et dispose d'outils qu'il peut manipuler cf. fig. 3. Chaque agent virtuel dispose d'un véhicule. Ce véhicule est représenté sur l'écran par trois icônes attachées à trois modes de déplacement : translation horizontale, verticale et rotation.

Conceptuellement, ce système peut se décrire comme une mémoire partagée par un ensemble de processus interagissant sur celle-ci. Cette mémoire partagée se répartit en mondes. Chaque monde représente un ensemble spécifique d'objets et de paramètres complètement distincts des autres mondes. Un processus de DIVE est un membre d'exactly un seul monde. Il existe des processus de visualisation et des processus de type application :

- Un processus de visualisation a pour tâche d'afficher périodiquement (environ tous les dixièmes de secondes suivant le matériel) le monde dont il est membre.
- Un processus de type application agit sur le monde en modifiant les objets et en le signalant aux autres processus afin de garder un monde cohérent.

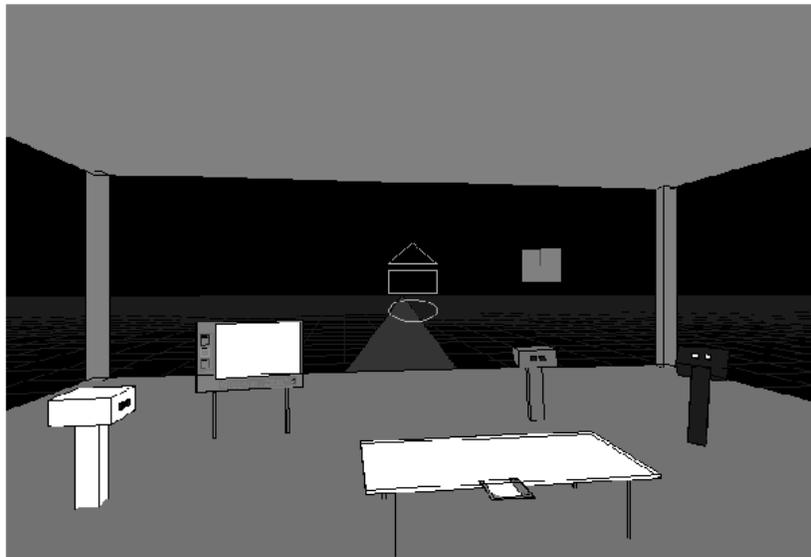


fig. 3 Le monde Conférence de DIVE vu par un des quatre utilisateurs connectés.

4. *Corpus de dialogues*

Nous avons réuni un corpus de dialogues afin d'étudier l'utilisation de la langue orale dans ces environnements virtuels. Cette démarche est cruciale pour identifier les problèmes n'apparaissant pas dans la réalité et laisser de côté ceux ne semblant intéressants que dans l'abstrait (Allen, 1994). Pour chaque dialogue une personne joue le rôle de l'utilisateur et une autre celui de l'agent virtuel. Ces dialogues sont enregistrés sur bandes magnétiques puis retranscrits sur machine. Nous avons collecté quatre dialogues, d1, d2, d3 et d4, avec à chaque fois un utilisateur différent, ceci pour obtenir une plus grande couverture linguistique. En revanche, le rôle de l'agent virtuel est toujours interprété par la même personne afin de garder une cohérence globale dans son discours. Des extraits de ces dialogues sont représentés sur les fig. 4 et fig. 5. U est l'utilisateur et A est l'agent virtuel.

4.1 *Extraits du corpus*

A	<i>nous sommes connectés au monde robot.</i>
U8	<i>tourne sur toi-même.</i>

A	<i>vers la droite ou vers la gauche?</i>
U9	<i>vers la droite.</i>
A	<i>voilà.</i>
U10	<i>prends de la hauteur.</i>
U11	<i>arrête de monter.</i>
U12	<i>monte.</i>
A	<i>oui.</i>
U13	<i>stop.</i>

fig. 4 Extrait 1 du dialogue d4

U45	<i>va jusque là.</i>
A	<i>je me dirige vers la montagne.</i>
U46	<i>fait le tour de la montagne.</i>
A	<i>oui.</i>
U47	<i>retourne sur l'île précédente</i>
A	<i>je ne connais pas l'île précédente.</i>
U48	<i>regarde à droite.</i>
A	<i>voilà.</i>
U49	<i>encore.</i>
U50	<i>c'est ici.</i>
A	<i>je me dirige vers la montagne.</i>

fig. 5 Extrait 2 du dialogue d4

4.2. Statistiques

Nous avons calculé diverses statistiques sur ce corpus. Nous avons étudié chaque dialogue (corpus d1, d2, d3 et d4) un à un, puis le cumul de ces dialogues. Chacune de ces études traite le discours de l'utilisateur, puis le discours de l'agent virtuel.

Nous parvenons aux constatations suivantes pour le discours de l'utilisateur :

- Le nombre de mots distincts est faible – le lexique est petit –.
- La croissance des mots distincts tend à se stabiliser malgré des utilisateurs différents.
- Le nombre de mots uniques est très important. Il y a 48% de mots distincts.
- La taille moyenne des énoncés est faible : 5 mots et au maximum 19.
- Les utilisateurs s'expriment principalement par ordres (d3 et d4) ou par ordres et questions (d1 et d2). Les utilisateurs de d1 et d2 ne connaissaient pas le système contrairement à ceux de d3 et d4. Cette différence concernant le nombre de questions peut s'expliquer ainsi.
- Les utilisateurs s'expriment rarement par négations : 2%.

5. Prototype d'Interface de dialogue oral dans les mondes virtuels

5.1. Architecture

Le système se compose notamment d'une carte de reconnaissance vocale, d'un analyseur syntaxique, d'un analyseur sémantique, d'un gestionnaire de dialogue et d'un gestionnaire d'actions. Ces modules s'articulent dans l'architecture de la fig. 6.

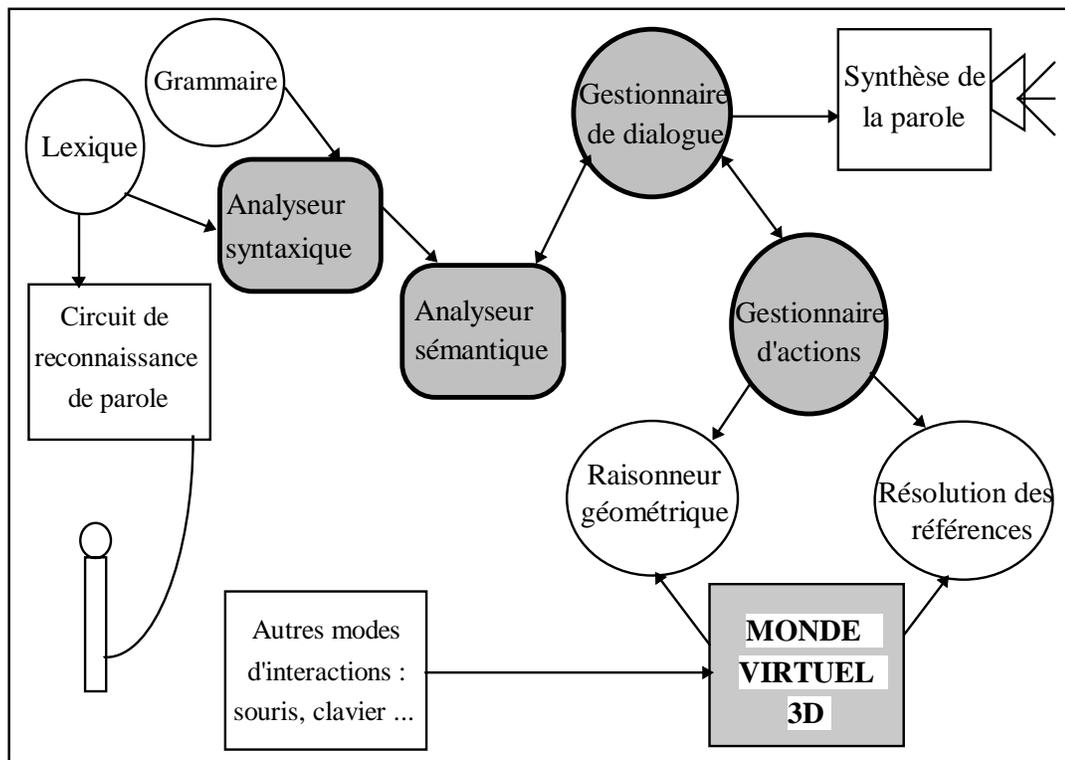


fig. 6 Architecture.

5.2. Analyse syntaxique

5.2.1 Analyse du corpus : création du lexique de l'utilisateur

L'analyseur syntaxique accepte des mots provenant du système de reconnaissance vocale. Il utilise un lexique fondé sur l'analyse des corpus. Il consiste dans l'ensemble des 327 mots distincts de l'utilisateur que nous avons catégorisé par des catégories lexicales. Chaque entrée du lexique comprend en outre son codage phonétique, son orthographe et une liste de traits de sous-catégorisation. Par exemple, le codage de *le*, *la*, *table*, *avance* est montré par la fig. 7.

Mots	Traits
le	DET le DET nombre = sing DET genre = masc
la	DET la DET nombre = sing DET genre = fem
table	NOUN table NOUN nombre = sing NOUN genre = masc
avance	VERB avance VERB nombre = sing

fig. 7 Un exemple de lexique.

L'analyse des énoncés de l'utilisateur fait ressortir l'utilisation de locutions prépositives et adverbiales. Ces groupes de mots ont à eux seuls une fonction grammaticale. Ils sont traités par notre analyseur comme une seule entité linguistique. Ces locutions correspondent à des entrées supplémentaires dans le lexique dont les principales sont représentées par la fig. 8.

Locutions prépositives		Locutions adverbiales	
à l'intérieur de	à_l_intérieur_de	à droite	à_droite
du côté de	du_côté_de	à gauche	à_gauche
en arrière de	en_arrière_de	en arrière	en_arrière
en avant de	en_avant_de	en avant	en_avant
en face de	en_face_de	en bas	en_bas
jusqu'à	jusqu'à	en face	en_face
		en haut	en_haut

fig. 8 Quelques locutions prépositives et adverbiales.

5.2.2 Analyse syntaxique

Nous avons développé un analyseur syntaxique fondé sur les techniques de Chart (El Guedj, 1994). Cet analyseur accepte des treillis de mots permettant le traitement en parallèle de plusieurs hypothèses lexicales. Il utilise des grammaires syntagmatiques et de dépendances enrichies d'équations d'unification. Pour cette application, nous nous sommes restreints à une grammaire syntagmatique du fait de la brièveté des énoncés à analyser – 5 mots en moyenne. L'analyseur accepte les mots séquentiellement et les analyse au fur et à mesure de leur énonciation jusqu'à ce que la phrase soit complète. Il peut opérer en mode descendant ou en mode ascendant. L'analyseur rejette tout mot entraînant une phrase agrammaticale, c'est à dire lorsque aucune règle syntaxique ne peut s'appliquer pour traiter ce mot. Un exemple de grammaire syntagmatique avec équations d'unification est décrit en fig. 9.

Règles syntagmatiques	Équations d'unification
PH -> PHRASE FIN	
PHRASE -> GN GV	PHRASE type = affirmation
	GN nombre = GV nombre
PHRASE -> GV	PHRASE type = ordre
GN -> DET NOM	DET nombre = NOM nombre
	DET genre = NOM genre
	GN nombre = DET nombre
	GN genre = NOM genre
GV -> VERBE	GV nombre = VERBE nombre

fig. 9 Un exemple de grammaire syntagmatique avec équations d'unification.

Notre système utilise une grammaire permettant d'analyser syntaxiquement tous les ordres de l'utilisateur présents dans le corpus. La fig. 10 représente cette grammaire. Les autres groupes syntagmatiques sont classiques et correspondent à la fig. 11. Les catégories lexicales apparaissant dans ces deux figures sont décrites par la fig. 12.

Groupes de règles d'analyse des ordres	Rôles des règles
R_ORDRE -> R_ORDRE0	Analyse récursive d'un ordre ou d'un ensemble d'ordres.
R_ORDRE -> R_ORDRE0 R_ORDRE	
R_ORDRE -> R_ORDRE0 CONJCOORD R_ORDRE	
R_ORDRE0 -> R_VIMP	Analyse d'un ordre simple qui consiste en un groupe verbal à l'impératif et un ou plusieurs autres groupes syntagmatiques.
R_ORDRE0 -> R_VIMP R_GN	
R_ORDRE0 -> R_VIMP R_GPREPN	
R_ORDRE0 -> R_VIMP R_GPREPV	
R_ORDRE0 -> R_VIMP VINF R_GPREPN	
R_VIMP -> V	Analyse du groupe verbal.
R_VIMP -> V R_ADV	
R_VIMP -> V R_PRONOM	
R_VIMP -> V R_PRONOM R_ADV	

fig. 10 Règles d'analyse des ordres.

Autres règles d'analyse syntagmatique	Rôles des règles
R_GN	analyse d'un groupe nominal.
R_GPREPN	analyse d'un groupe prépositionnel nominal.
R_GPREPV	analyse d'un groupe prépositionnel verbal.
R_ADV	analyse d'une suite d'adverbes.
R_PRONOM	analyse d'un pronom personnel, ou adverbial.

fig. 11 Règles d'analyse de syntagmes.

Quelques catégories lexicales	Rôles des catégories
VINF	catégorie du lexique des verbes à l'infinitif.
V	catégorie du lexique des verbes conjugués (les équations d'unifications de R_VIMP vérifient l'emploi de l'impératif).
CONJCOORD	catégorie du lexique des conjonctions de coordinations.

fig. 12 Quelques catégories lexicales.

5.3 Analyse sémantique et discursive

Dans ce prototype, les énoncés de l'utilisateur sont pour la plupart des énoncés relatifs à des actions : des ordres. Un énoncé de ce type est constitué d'une ou de plusieurs actions.

- Exemple 1 : Un énoncé composé d'une seule action :
U: Monte sur la table blanche.
-> action1 = monter sur OBJET
- Exemple 2 : Un énoncé composé de trois actions :
U: avance en montant et regarde en bas.
-> action1 = avancer
-> action2 = monter
-> action3 = regarder DIRECTION

L'analyseur sémantique décompose le Chart, retourné par l'analyseur syntaxique, en une liste d'actions. D'après les énoncés qui apparaissent dans notre corpus, cette décomposition dépend de plusieurs paramètres :

- le nombre de propositions et leur fonction dans l'énoncé. L'énoncé de l'exemple 2 ci-dessus comporte deux propositions indépendantes
- le nombre de verbes dans chaque proposition et la fonction de chacun de ces verbes : s'il dirige la proposition principale, s'il appartient à un complément de circonstance...

Nous utilisons une grammaire de cas afin de représenter sémantiquement une action sous la forme de structure prédicative (Mast, 1993). Le verbe de l'action correspond au nom du prédicat. Les autres constituants de la phrase sont associés aux paramètres de la structure prédicative. Par exemple, le verbe « monter » et le verbe « regarder » sont associés aux prédicats:

```
monter([sens, ALLER],[lieu_par_defaut, EN_HAUT]).
regarder([sens,REGARDER],[lieu_par_defaut, EN_AVANT]).
```

La liste des actions est ensuite transmise au gestionnaire d'actions. Si ces actions sont exécutables, le gestionnaire du dialogue acquiesce l'ordre de l'utilisateur par un message positif aléatoire. Dans le cas contraire le gestionnaire de dialogue signale oralement la cause de la non-exécution de l'ordre. La synthèse orale permet à l'utilisateur de modérer son attention visuelle ordinairement accaparée par le flot continu d'images.

Le gestionnaire d'actions vérifie et exécute les actions transmises par l'analyseur sémantique. Il utilise un résolveur de références et un module de raisonnement géométrique.

5.4 La résolution des références

5.4.1 Le nommage et le codage des objets

Associer un nom à un objet est parfois une tâche complexe. Les utilisateurs peuvent utiliser un vocabulaire différent pour désigner la même chose. Ils peuvent aussi considérer certains objets comme des compositions ou des hiérarchies alors que dans la base de données, ils forment des entités uniques. Une maison peut se représenter comme un ensemble de lignes polygonales ou bien un ensemble de sous objets tels qu'un toit, une porte, des fenêtres, etc. Ces sous objets étant constitués soit d'autres sous structures, soit des lignes polygonales. De plus, il faut tenir compte de l'orientation des objets : le « devant d'une maison » n'est pas la même chose que le « devant d'un cube ».

Dans le prototype actuel (Revolta-Blaudeau, 1995), les objets sont des entités géométriques de la base de données du monde auxquelles on associe un concept par exemple *maison*. Ce nommage est bien sûr dépendant de la structure de la base de données géométrique. Nous avons cherché à établir les correspondances les plus pertinentes en fonction de notre corpus.

Pour ce qui concerne l'orientation, nous avons cherché à respecter un principe unique : si on peut aller à l'intérieur de l'objet, le centre de son référentiel est au centre de l'objet et c'est l'endroit où sera placé l'utilisateur s'il demande à rentrer à l'intérieur. Il faut donc le définir judicieusement. De plus, pour repérer sa face de devant s'il en a une, l'axe des z de son référentiel est orienté vers cette face.

5.4.2 L'algorithme de résolution de références

Les références et les désignations ambiguës sont multiples dans notre corpus. Elles proviennent notamment de références déictiques

va ici

ou de possibilité de choix multiples

dirige toi vers la maison (parmi plusieurs)

Pour lever ces ambiguïtés, nous avons utilisé deux critères : le fait qu'un objet soit dans le champ de vision de l'utilisateur et un focus dont on a pourvu chaque objet. Ce focus s'inspire de (Huls, 1995; Karlgren, 1995). Il consiste en un coefficient mis à jour en fonction des interactions de l'utilisateur. Il permet de résoudre d'une manière simple les ambiguïtés de désignation. Pour chaque utilisateur, chaque objet du monde dispose d'un focus – un nombre entier –. Le focus d'un objet devient majorant de tous les autres lorsque l'utilisateur clique sur cet objet (désignation sans équivoque) ou lorsqu'il le désigne dans une phrase.

5.5. Le module de raisonnement géométrique

Le module de raisonnement géométrique intervient comme intermédiaire entre les énoncés de l'utilisateur et le monde virtuel. Il constitue la principale entité de raisonnement de l'agent. Nous présentons ici le raisonnement que cet agent met en œuvre lorsque l'utilisateur lui donne des ordres de navigation. Ces ordres consistent dans des énoncés qui contiennent toujours un verbe. Pour l'instant, nous distinguons cinq catégories de verbes. Ils correspondent à la catégorie *Changement de lieu* cité par (Aurnague, 1994) :

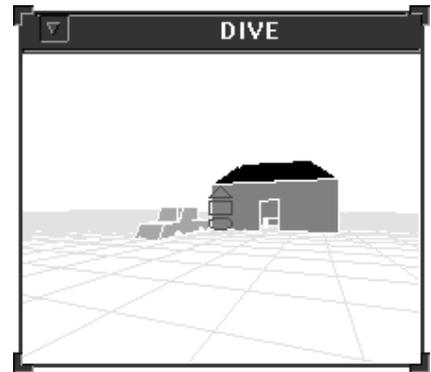
- « aller » (*va, avance, entre, monte, sors, ressorts...*) correspond à un changement de lieu avec éventuellement une rotation du personnage,
- « retourner » (*retourne, reviens*). Cette classe est distincte de la précédente par le fait que l'objet devra être recherché par le focus et non par la visibilité,
- « arrêter », (*arrête, stop*)
- « continuer » (*continue*)
- « tourner » (*tourne, oblique, regarde*) qui correspondra à une simple rotation.

La fig. 13 présente un exemple de dialogue réalisé par notre prototype ainsi qu'une série de copie d'écran associée.

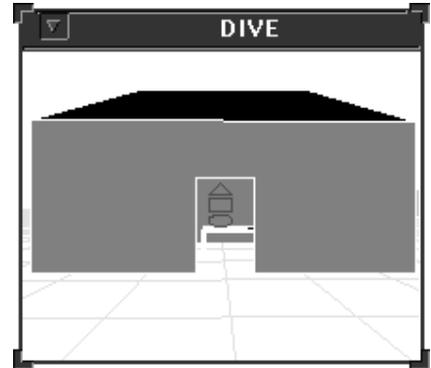
Énoncés et interactions de l'utilisateur	Énoncés de l'agent	Copies d'écran
--	--------------------	----------------

Bonjour bienvenue dans le monde Ithaques

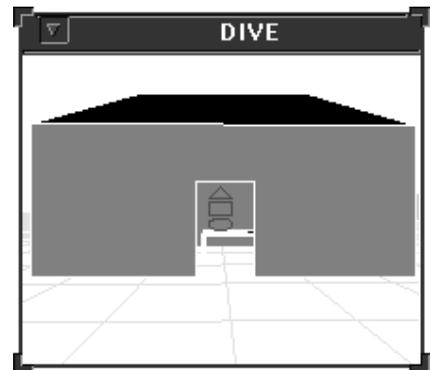
va devant la maison.



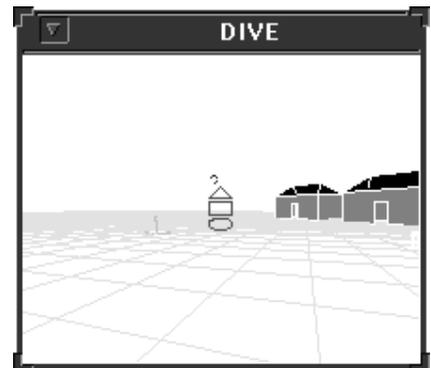
Voilà.



Regarde derrière toi.



Voilà.



*Est-ce que tu peux entrer
dans la maison ?*

Il y en a plusieurs.

*-- L'utilisateur clique sur le
bouton gauche de la souris --*

entre dedans.

D'accord.

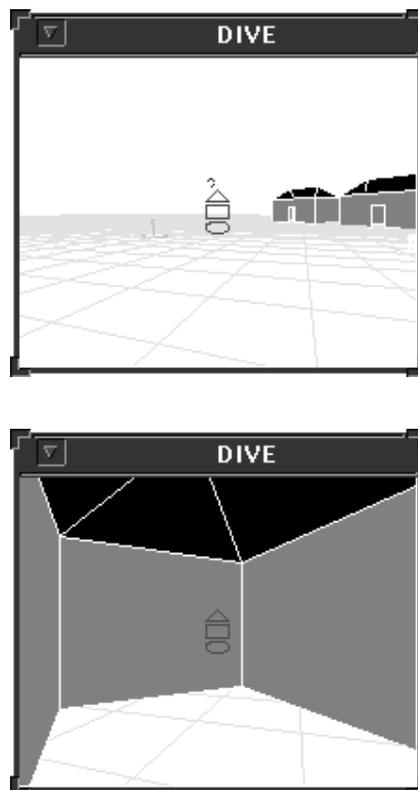


fig. 13 Exemple de dialogue.

6. Conclusion

Nous avons réalisé l'intégration d'un prototype d'agent conversationnel pour naviguer oralement dans un système de réalité virtuelle. Dans un premier temps, nous avons justifié l'utilisation d'un tel type d'agent dans un monde virtuel. Puis, nous avons présenté un corpus de dialogue et les statistiques que nous avons obtenues à partir de celui-ci. Ensuite, nous avons détaillé l'architecture du système et ses différents composants. Enfin, nous avons décrit l'implantation actuelle de notre agent.

Ce type d'agent conversationnel offre des perspectives importantes. Nous avons notamment le projet de l'adapter à la navigation orale dans une synthèse virtuelle de la ville de Rome antique. Cette synthèse étant réalisée par le Centre d'Études et de Recherches pour l'Antiquité de l'Université de Caen. Par ailleurs, au fur et à mesure de nos tests nous mettons en évidence de nouveaux besoins, tels que de nouvelles actions de navigation ou la création d'un agent de manipulation d'objets. Plus généralement nous pensons que ce type d'agent offre un cadre d'expérimentation unique pour le dialogue, la linguistique informatique ou encore le raisonnement géométrique.

Références

- ALLEN, J.F., SCHUBERT, L.K., FERGUSON, G., HEEMAN, P., HEE HWANG, C., KATO, T., LIGHT, M., MARTIN, N.G., MILLER, B.W., POESIO, M. AND TRAUM, D.R.: The TRAINS Project: A case study in building a conversational planning agent, TRAINS Technical Note 94-3, University of Rochester, New York, September 1994.
- ANDERSSON, M., CARLSSON, C., HAGSAND, O. AND STÅHL, O.: DIVE, The Distributed Interactive Virtual Environment, Technical Reference, Swedish Institute of Computer Science, Kista, Suède, March 1994.
- AURNAGUE, M., SABLAYROLLES, P., JAYEZ, J.: Les informations spatio-temporelles dans les constats d'accidents: Représentation du contenu sémantique et raisonnement, Traitement Automatique des Langues, vol. 35, n° 1, pp. 107-130, 1994.
- BALL, G. ET AL: Likelike Computer Characters: The Persona Project at Microsoft Research, in Software Agents, J. Bradshaw ed., MIT Press, A paraître.

- BENFORD, S. AND FALHEN, L.:** A Spatial Model of Interaction in Virtual Environments, Proceedings of the 3rd European Conference on CSCW, Kluwer Academic Press, Dordrecht, September 1993.
- BOLT, R.A.:** Put That There: Voice and Gesture at the Graphic Interface, Computer Graphics, vol. 14, n° 3, pp. 262-270, 1980.
- COTECH:** Minutes of the COTECH Workgroup: Virtual and Augmented Environments for CSCW, Department of Computer Science, University of Nottingham, Nottingham, Angleterre, 1995.
- EL GUEJ, P.O. ET NUGUES, P.:** A chart parser to analyze large medical corpora, Proceedings of the 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Baltimore, pp. 1404-1405, November 1994.
- GODEREAUX, C., DIEBEL, K., EL-GUEJ, P.O., REVOLTA, F. ET NUGUES, P.:** Interactive Spoken Dialogue Interface in Virtual Worlds, One-Day Conference on Linguistic Concepts and Methods in Computer-Supported Cooperative Work, London, November 1994, Actes à paraître chez Springer Verlag.
- GRUDIN, J.:** Computer-Supported Cooperative Work: History and Focus, Computer, vol. 27, n° 5, pp. 19-26, May 1994.
- HULS, C., BOS, E. AND CLAASSEN, W.:** Automatic Referent Resolution of Deictic and Anaphoric Expressions, Computational Linguistic, vol. 21, n° 1, pp. 59-79, 1995.
- KARLGREN, J., BRETAN, I., FROST, N. AND JONSSON, L.:** Interaction Models, Reference, and Interactivity in Speech Interfaces to Virtual Environments, 2nd Eurographics Workshop, Monte Carlo, Darmstadt, Fraunhofer IGD, 1995.
- MAST, M., KUMMERT, F., EHRlich, U., FINK, G.A., KUHN, T., NIEMANN, H. AND SAGERER, G.:** A speech understanding and dialog system with a homogeneous linguistic knowledge base, IEEE Transactions on pattern analysis and machine intelligence, vol. 16, n° 2, 179-194, 1994.
- NUGUES, P., GODEREAUX, C., EL GUEJ, P.O. AND CAZENAVE, F.:** Question answering in an Oral Dialogue System, In: Proceedings of the 15th Annual International Conference IEEE/Engineering in Medicine and Biology Society, Paris, vol. 2, pp. 590-591, 1993.
- NUGUES, P., CAZENAVE, F., EL GUEJ, P.O. AND GODEREAUX, C.:** Un système de dialogue oral guidé pour la génération de comptes rendus médicaux, In: Actes du 9e congrès de l'AFCEC-INRIA Reconnaissance de Formes et Intelligence artificielle, Paris, vol. 2, pp. 79-88, janvier 1994
- QUEAU, P.:** Le virtuel: vertus et vertige, Champ Vallon/INA, Seyssel, 1993.
- REVOLTA-BLAUDEAU, F.:** Navigation dans les mondes virtuels. Résolution des références, raisonnement géométrique et exécution d'actions, Mémoire de Diplôme d'études approfondies, Université de Caen, Septembre 1995.
- ROBERTSON, C.G., CORD, S.K. AND MACKINLAY, J.D.:** Information visualization using 3D interactive animation, Communications of the ACM, vol. 36, n° 4, pp. 57-71, 1993.

Adresse des auteurs: Christophe Godéreaux, Pierre-Olivier El Guedj, Frédéric Revolta et Pierre Nugues, GREYC, ISMRA et Université de Caen, 6, Boulevard du Maréchal Juin, F-14050 Caen.

A Conversational Agent to Navigate in Virtual Worlds (Summary)

We describe the prototype of an interactive spoken dialogue interface in a virtual reality system. This prototype accepts utterances from a user enabling him or her to navigate into relatively complex virtual worlds. We first justify the significance of this type of interface in the communication quality between a user and virtual worlds. Then we present a corpus of dialogues and statistical results. Next we describe our prototype which includes a speech recognition device together with a speech synthesizer. The dialogue prototype consists in a syntactic chart parser for spoken words; a semantic analyzer handling the meaning of utterances; a reference resolution system; and a dialogue analyzer. It is integrated in the DIVE virtual reality environment developed at the Swedish Institute of Computer Science.