

ANALYSE PRAGMATIQUE D'UNE INTERACTION

Ozkan Nadine et Caelen Jean
ozkan@icp.imag.fr caelen@icp.imag.fr

ICP / INPG
46 avenue Félix Viallet
38031 Grenoble cedex
76.51.45.33

Résumé

Cet article présente une étude pragmatique de la structure de la communication entre deux agents autour d'une tâche bien circonscrite. Cette étude s'inscrit dans une démarche de recherche dont l'objectif est l'élaboration d'un modèle de dialogue homme-machine qui soit efficace et adaptatif. Au plan théorique, la communication est vue comme un processus de construction d'un espace commun nécessitant une adaptation mutuelle des interlocuteurs. Au plan méthodologique, nous procédons empiriquement - c'est-à-dire avec une démarche ascendante partant des données-terrain vers un modèle. Nous avons donc effectué une expérience impliquant deux agents (humains) coopérant autour de tâches données. Le domaine des tâches est la construction de figures spatiales. Nous avons recueilli, formalisé et analysé les échanges verbaux des deux agents. Cette analyse est centrée sur leurs stratégies communicationnelles de désignation et de positionnement des objets dans l'espace. Elle permet de repérer la structure des échecs communicationnels, les causes de ces échecs et d'en tirer des conclusions pour la communication homme-machine.

Mots clé : interaction homme-machine, pragmatique, représentation des connaissances

Summary

We present a pragmatic study of the structure of communication between two agents around a well-defined task. This study constitutes one of the research stages towards the development of a model of human-computer interaction. Theoretically, communication is defined as the process of construction of a common space requiring mutual adaptation from the participants. Methodologically, we proceed empirically, from data to model. We designed an experiment involving two (human) agents cooperating around given tasks. The tasks involved spatial construction of figures. The verbal exchanges between the agents were formalized so as to allow the analysis of the structure, causes and repairs of communicational failures. Conclusions are drawn as to the design of human-computer interfaces.

Key words : human-computer interaction, pragmatics, knowledge representation

Introduction : problématique et contexte

Les perspectives récentes sur la communication homme-machine la conçoivent comme une collaboration entre deux agents de nature différente. Dans le domaine de la communication humaine, les théories actuelles ont comme point focal l'adaptation mutuelle des interlocuteurs et leur activité commune de construction de signification au cours des échanges. L'hypothèse de nos travaux est que les certains concepts provenant des théories de la communication humaine (en particulier la communication langagière - voir la section du cadre théorique) peuvent être utiles pour la communication homme-machine. Elle se trouve justifiée par certains travaux [SUC87,WEI76] portant sur les attentes envers les ordinateurs en tant qu'acteurs sociaux et agents cohérents.

Par conséquent, le modèle de communication homme-machine que nous développons s'appuie sur l'analyse de la communication entre humains d'une part et entre humain et machine d'autre part. Notre stratégie de recherche est la suivante : pour une même tâche, soit la construction de figures spatiales, nous analysons des situations communicationnelles de plus en plus proches de notre cible, c'est-à-dire la communication homme-machine dirigée par la tâche. La première étape concerne la description de figures spatiales par un agent de telle façon que ces figures puissent éventuellement être reconstruites par un autre agent (hypothétique). Il s'agit donc là d'une situation de monologue, l'agent descripteur n'ayant aucun interlocuteur au moment de la description [MON91]. La seconde étape est la situation communicationnelle décrite dans cet article. Il s'agit de la construction de ces mêmes figures partagée entre deux agents : un agent-instructeur, qui donne des instructions verbales, et un agent-manipulateur, qui exécute ces instructions à l'ordinateur. Cette fois, il s'agit véritablement d'un dialogue autour d'une tâche avec retour immédiat, les deux agents étant face à l'écran et pouvant suivre la construction au fur et à mesure des instructions. La troisième étape est en cours de réalisation. Il s'agit d'échanges homme-machine multimodaux simulés par la technique du Magicien d'Oz. L'étape finale est le dialogue homme-machine réel, qui intégrera le modèle de communication développé aux étapes précédentes.

La communication en tant qu'action

Faisant suite aux travaux d'Austin et Searle [AUS62] [SEA69], nous nous appuyons sur un cadre pragmatique où un énoncé langagier est considéré comme une action - de par le changement qu'il entraîne dans le monde - et où sa signification repose sur les connaissances partagées par les interlocuteurs et sur l'intention que chacun a et prête à son partenaire.

Le point focal de notre analyse sont les ruptures de communication, que nous définissons comme le fossé manifeste entre les connaissances des interlocuteurs, où l'un suppose à tort que ses propres connaissances mises en jeu dans un énoncé sont partagées par son partenaire (par exemple, l'utilisation par l'un d'un mot inconnu de l'autre). Ces connaissances peuvent être de nature lexicale, syntaxique, sémantique ou pragmatique. Une rupture de communication donne lieu à une réparation, définie comme une phase de négociation du concept en jeu, où les interlocuteurs tentent de combler le fossé entre leurs connaissances (par exemple, expliquer le mot inconnu jusqu'à ce qu'il soit signifiant - on dit alors que la signification du mot est négociée). Soulignons qu'une rupture de communication se produit lorsqu'un des interlocuteurs met ce fossé en évidence en initiant une négociation. En effet, il se peut que le fossé ne soit jamais rendu explicite pour diverses raisons : la bonne marche de la communication n'est pas véritablement en jeu (les humains fonctionnent très bien en présence d'ambiguïté), le coût d'une négociation est jugé trop fort, le contexte social d'interaction se prête mal à une interruption du fil du discours, etc. La phase de négociation n'a lieu que suite à la détection du fossé par l'un des interlocuteurs, à son évaluation de l'apport possible de la négociation, et à sa décision d'indiquer le fossé à son partenaire.

À l'issue d'une phase de négociation, le fossé ayant été comblé, les concepts négociés entrent dans le lot des connaissances partagées par les interlocuteurs. Il est donc clair que la communication, qui repose sur les connaissances partagées, contribue aussi à étendre ces connaissances, via les mécanismes de rupture et de négociation. C'est dans ce sens qu'elle est une construction commune de signification. Nous considérons les ruptures comme des points importants d'articulation du dialogue en ce qu'elles révèlent les suppositions du locuteur sur les connaissances de son partenaire nécessaires à sa compréhension de l'énoncé.

La logique des mondes possibles

Quelle est la nature des connaissances en jeu dans la communication humaine ? Typiquement, ces connaissances appartiennent à des domaines distincts (connaissances de la langue, connaissances de "sens commun", connaissances du sujet discuté, connaissances de sujets connexes etc..) qui sont entremêlés au cours de la communication. Nous nommons ces domaines des mondes d'interaction, et définissons un monde comme un ensemble cohérent d'actions et d'objets liés sémantiquement et pragmatiquement. (Il est difficile de donner une définition plus précise, puisque les éléments constitutifs d'un monde et les liens entre eux dépendent avant tout de la situation.)

La communication homme-machine se situe également à la jonction de plusieurs mondes d'interaction. Contrairement à la communication humaine toutefois, et de par la

nature différente des deux agents interagissant, elle se situe a priori et indépendamment de toute considération de tâche ou de technologie à la jonction d'au moins deux mondes :

- le monde virtuel - celui de la tâche telle qu'elle est signifiante pour l'utilisateur et qui lui sert de guide ;
- le monde artificiel - celui qui est présenté par l'interface. Précisons que plusieurs niveaux de mondes artificiels peuvent être distingués : d'abord le monde métaphorique de haut niveau (métaphore du bureau avec dossiers, poubelle etc.) ; à un niveau intermédiaire, le monde des concepts informatiques tels que fichiers, données, applications etc. ; enfin, celui des concepts de plus bas niveau (registres de mémoire, pixels, bits etc.). Précisons également que le monde artificiel présenté par l'interface peut être un enchevêtrement de plusieurs de ces niveaux (par exemple le "Finder" du Macintosh, qui présente des concepts à la fois métaphoriques et informatiques).

Méthodologie

Nous avons analysé la nature des ruptures au cours d'un dialogue entre deux humains autour d'une tâche de construction de figures à l'ordinateur, dans le but d'appliquer ces résultats à la conception d'une interface personne-machine pour un logiciel de dessin architectural. La situation expérimentale était la suivante (voir figure 1) : une série de figures étaient présentées à l'agent-instructeur qui donnait verbalement des instructions à l'agent-manipulateur afin que ce dernier les exécute à l'ordinateur via la souris et le clavier. Les productions verbales des deux agents étaient enregistrées pour analyse.

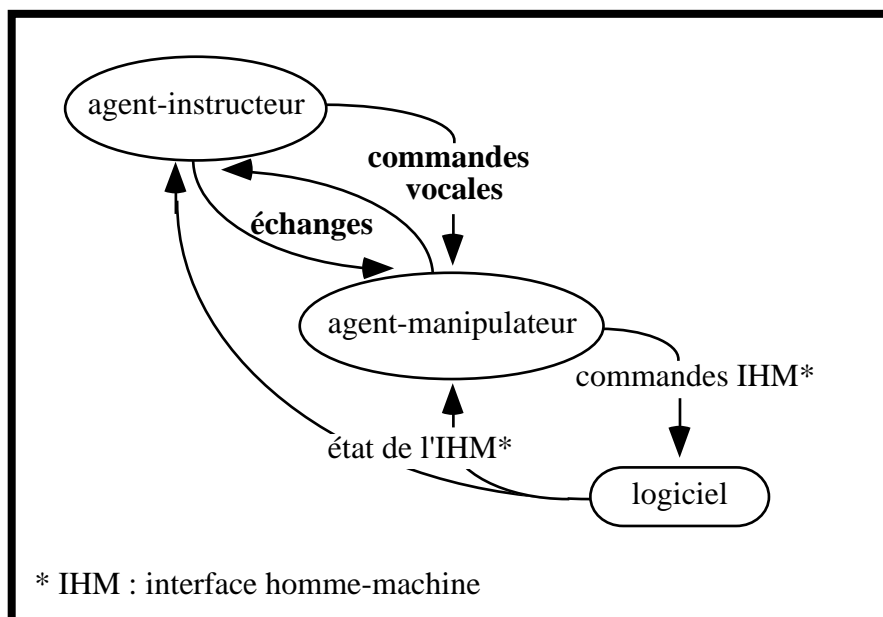


figure 1 - Echanges communicationnels de la situation expérimentale. (Les échanges en caractère gras sont l'objet de l'étude.)

Les figures données à l'agent-instructeur étaient réparties en trois groupes (figure 2) : un premier groupe constitué de figures abstraites (figure 2a), un second groupe constitué de figures structurellement identiques à celles du premier groupe, mais représentant des pièces reliées par des portes (l'icône en forme de losange - figure 2b.), et un troisième groupe constitué toujours des mêmes figures, mais contenant également des icônes figurant des meubles et fenêtres (figure 2c). Ainsi, chaque groupe de figures est de plus en plus concret ou représentationnel. Nous considérons chaque tâche de description d'un groupe de figures comme un monde distinct tel que défini plus haut. L'expérience met donc en jeu trois mondes de façon successive.

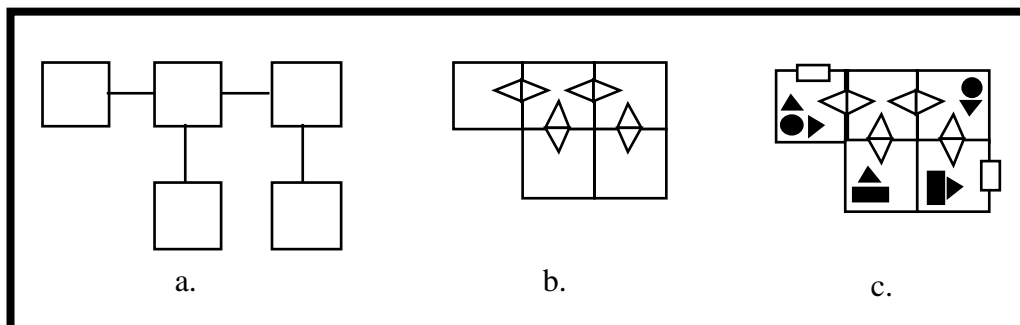


figure 2 - Les trois types de tâches

Huit agents-instructeurs et un seul agent-manipulateur ont pris part à cette expérience. Parmi les huit agents-instructeurs, six sont des hommes et deux sont des femmes. Sept sont ingénieurs. L'analyse de leurs échanges avec l'agent-manipulateur porte sur leurs stratégies communicationnelles de désignation et de positionnement des objets dans l'espace telles qu'elles sont comprises par l'agent-manipulateur. Rappelons que la notion de rupture s'entend ici au sens communicationnel défini plus haut. Notre analyse ne porte pas sur les erreurs au plan de la tâche. Rappelons également que les ruptures conversationnelles peuvent se situer aux plans lexical, syntaxique, sémantique ou pragmatique. Nous ne nous intéressons ici qu'aux ruptures pragmatiques.

Formalisation, observation et analyse du corpus

Afin de découvrir comment se constituent les connaissances communes et quand et pourquoi se produisent les ruptures, nous avons formalisé les connaissances impliquées dans les trois mondes des tâches. Notre propos ici étant de montrer l'utilité de ces représentations formelles pour le domaine de la conception d'interface et non pour le domaine de l'intelligence artificielle, nous nous contenterons de les décrire très brièvement.

La construction de figures se fait par une succession de positionnements d'objets relativement à d'autres objets déjà placés. Nous avons représenté les objets par des frames et leur position relative par des prédicats. Les frames sont reliés en hiérarchie de façon classique par des liens de type "est-un" ou "type-de" et ont des attributs propres ou hérités [MIN75]. Les frames sont également groupés selon leur appartenance au même monde. Chaque monde est donc constitué d'objets représentés par des frames distincts. La position relative d'un objet telle que décrite par l'agent-instructeur est représentée par un prédicat. Ce prédicat spécifie l'objet défini, l'objet de référence ainsi que la relation spatiale entre eux.

Les productions verbales des deux agents sont analysées comme suit : tout d'abord, elles sont segmentées en actes de langage. (Rappelons d'un acte de langage est l'unité élémentaire de communication.) Les objets auxquels l'acte de langage fait référence (explicitement ou implicitement) sont ensuite représentés sous forme de frames et les relations spatiales (explicites ou implicites), sous forme de prédicat. Cette représentation formelle de chaque acte de langage permet le repérage rigoureux des éléments implicites au discours mais néanmoins nécessaires à son interprétation. Elle permet également la détection des causes des ruptures et des négociations, ainsi que de leur évolution au cours des échanges.

Le fossé entre les connaissances des agents se manifeste de quatre façons :

- 1) la confirmation de l'exécution d'une instruction par l'agent-instructeur (exemple : "oui, c'est bien ça."). Une négociation se termine toujours par la confirmation de part et d'autre qu'une entente a été atteinte. Dans ce cas, il n'y a pas de négociation mais il y a confirmation. L'agent-instructeur estime que son énoncé (celui qui précède la confirmation) aurait pu potentiellement causer une rupture et que cette rupture a été évitée par l'interprétation correcte de l'agent-manipulateur, d'où la confirmation.
- 2) la reformulation ou la redite d'une instruction précédente de manière plus explicite ou complète sans sollicitation de la part de l'agent-manipulateur (exemple : "par ... je veux dire ..."). Comme au premier cas, l'agent-instructeur estime que son énoncé (celui qui précède la reformulation) peut potentiellement causer une rupture. Ici, il prend toutefois lui-même l'initiative d'éviter la rupture en le reformulant.

Dans ces deux cas, l'agent-instructeur est manifestement conscient du fossé et de la rupture potentielle causée par l'instruction précédant la confirmation ou la reformulation. Pourquoi alors décide-t-il d'énoncer une instruction qu'il juge problématique? Cette instruction a deux fonctions : sa fonction actionnelle proprement dite (accomplissement d'une étape de la tâche) et une fonction communicationnelle qui

est d'informer l'agent-manipulateur d'une formulation problématique (par exemple ambiguë ou incomplète) mais préférable (par exemple moins coûteuse), et qui sera reprise par la suite. Dans le premier cas, cette fonction communicationnelle est implicite (ce qui met en péril la fonction actionnelle) : malgré la possibilité de rupture, l'agent-instructeur compte sur la faculté d'interprétation de l'agent-manipulateur. Dans le second cas, du fait de la reformulation, la fonction communicationnelle est explicite (et la fonction actionnelle n'est pas problématique). Il s'agit ici véritablement de ce processus de construction d'un espace commun de signification initié par l'agent-instructeur afin d'améliorer l'efficacité de sa communication avec l'agent manipulateur.

- 3) la correction par l'agent-instructeur d'une erreur d'interprétation de la part de l'agent-manipulateur. Celle-ci se manifeste par son action sur l'interface dont le résultat n'est pas celui attendu par l'agent-instructeur (exemple : "non, je voulais dire ...").
- 4) la demande de clarification de la part de l'agent-manipulateur, suite à une instruction incomplète ou ambiguë (exemple : "tu n'as pas spécifié ...").

Seuls ces deux derniers cas donnent lieu à des ruptures et à des négociations (négociations initiées respectivement par l'agent-instructeur et par l'agent-manipulateur). Ces ruptures se produisent dans deux types de situations : Tout d'abord lorsque de l'information répétitive est implicite dans l'acte de discours (par exemple, lorsque l'agent-instructeur oublie de dire que le premier élément de chaque figure doit être au bas de l'écran). L'agent-instructeur suppose que l'agent-manipulateur reconnaîtra une similitude entre la situation courante et une situation passée et qu'il retrouvera les éléments manquants à partir de cette similitude. La seconde situation se produit lorsque l'agent-instructeur opère dans le cadre d'un monde spécifique et change de monde de façon abrupte et inattendue. (Par exemple, au cours de la tâche 3 un agent-instructeur dit : "mets une porte sur le côté gauche du carré". Ceci est un acte de langage mixte en ce sens que l'objet "porte" appartient au monde virtuel de la tâche 3 tandis que les objets "côté" et "carré" appartiennent au monde artificiel, celui de l'interface. Un acte de langage cohérent aurait été : "mets une porte sur le mur gauche de la pièce" ou encore "mets un losange sur le côté gauche du carré".) Ainsi, les mondes virtuel et artificiel sont simultanément présents à l'esprit de l'agent-instructeur, de sorte qu'il passe indifféremment de l'un à l'autre. De tels changements de mondes se produisent au sein d'un même acte de langage, tel qu'illustré par l'exemple, ainsi qu'entre deux actes de langage. Il ne s'agit pas ici de simple substitution lexicale ("côté" pour "mur" ou "carré" pour "pièce") mais bien de rupture pragmatique, car les objets de mondes différents ont des caractéristiques différentes et obéissent à des contraintes différentes, ce qui conditionne leur usage et leur signification. (Par exemple, les pièces faisant partie d'un plan architectural - monde virtuel - ne sont pas interpénétrables alors que les carrés d'un

dessin McDraw - monde artificiel - n'obéissent pas cette contrainte.) La violation d'une contrainte ou la contradiction d'une caractéristique d'un objet lors d'un tel changement de monde donne lieu à une rupture pragmatique.

Conséquences pour la conception d'IHM

Nous présentons ici les implications de notre analyse pour la conception d'interface homme-machine et plus particulièrement pour l'élaboration d'un modèle de dialogue. La stratégie d'interaction d'une interface homme-machine devrait permettre d'éviter le plus grand nombre de ruptures possible, tout en veillant à ce qu'aucune contrainte pragmatique ne soit violée. Nous avons observé que la détection d'un fossé entre les connaissances ne donne pas toujours lieu à une négociation (cas 1 et 2 ci-dessus), selon que cette dernière est jugée pertinente ou non. Une stratégie d'évaluation de la pertinence d'une rupture devrait, par conséquent être incorporée à l'interface homme-machine, afin d'assurer un dialogue opportun.

Nous avons également observé deux situations de ruptures s'articulant autour du processus de partage de connaissances (cas 3 et 4 ci-dessus) : dans le premier cas, l'agent-instructeur suppose que l'information répétitive est partagée tacitement par l'agent-manipulateur. Dans le domaine de la communication homme-machine, la négociation peut être évitée par la création et la mise-à-jour dynamique de l'historique des activités de l'utilisateur. Une interface homme-machine pourrait alors raisonner sur cet historique afin d'effectuer de la reconnaissance de plans et de d'inférer ainsi l'intention de l'utilisateur, pour tenter d'interpréter une instruction incomplète.

Dans le second cas, celui des changements de mondes, la rupture peut être évitée si l'action de l'utilisateur respecte les contraintes pragmatiques des mondes virtuel et artificiel. Ceci nécessite trois modules d'interface :

- la représentation explicite du monde artificiel ;
- la création dynamique (en cours de dialogue) d'une représentation du monde virtuel ;
- la création dynamique de liens sémantiques et pragmatiques entre les concepts du monde artificiel et ceux du monde virtuel.

Pour interpréter une instruction contenant un concept absent du monde courant d'interaction, l'interface pourrait alors parcourir les liens sémantiques ou pragmatiques entre les mondes et retrouver le concept en question. La mise-à-jour du monde virtuel de l'utilisateur devra se faire dynamiquement et exiger un minimum d'interruption du fil d'activité.

La figure 3 se veut une synthèse de la structure de la communication homme-machine telle qu'impliquée par les conclusions de l'étude présentées ci-dessus. Cette figure

présente un modèle local, c'est-à-dire où la détection d'un fossé communicationnel (par l'un ou l'autre des interlocuteurs) suit immédiatement l'action problématique. Il est bien entendu que ceci est un cas particulier d'un modèle plus global, où la rupture causée par un énoncé peut survenir après plusieurs tours de parole.

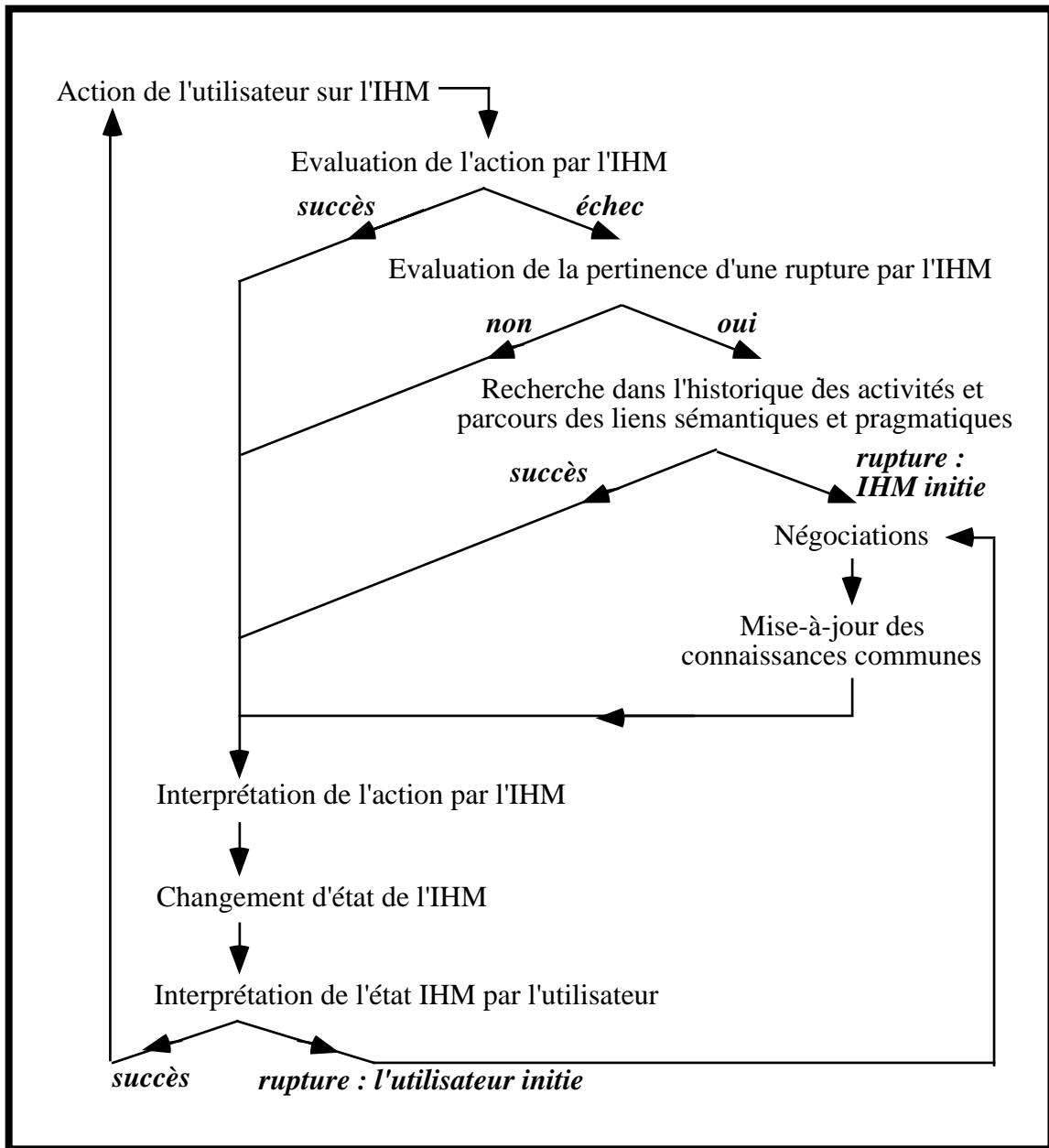


figure 3 - Structure de la communication homme-machine

Conclusions et perspectives

Deux facteurs conditionnent la bonne marche du processus de communication dans un dialogue dirigé par la tâche : la compréhension du but et du plan du partenaire et la cohérence de perception des mondes (monde virtuel de l'utilisateur pour l'interface et monde artificiel de l'interface pour l'utilisateur). Nous ne voulons pas toutefois impliquer par là que la réduction du nombre de ruptures devrait être un objectif communicationnel ni qu'elle constitue une mesure du succès communicationnel : d'une part comme nous l'avons vu, le déclenchement d'une rupture est une question d'évaluation de la situation ; d'autre part, les négociations constituent justement la manifestation explicite du processus d'élaboration commune de signification propre à la communication. Plutôt qu'être un empêchement à la communication, les ruptures et les négociations, si elles sont gérées de façon opportune, sont essentielles à son déroulement.

Il est évident que la mise en oeuvre des fonctionnalités proposées ci-dessus pour une interface monomodale (langagière) nécessite en amont des études des stratégies des utilisateurs, de l'évolution de ces stratégies et des liens qu'ils conçoivent entre concepts. Des études supplémentaires sont nécessaires pour aborder le dialogue multimodal. Dans l'expérience de communication orale décrite ici, la majorité des ruptures résulte de l'omission par l'agent-instructeur d'information répétitive concernant les relations spatiales entre objets (que nous avons représentée par des prédicats). Qu'en est-il d'une situation de communication multimodale (voix et geste), où plutôt que de décrire verbalement la position d'un objet à l'écran, l'utilisateur n'aura qu'à la désigner d'un geste ? Les causes des ruptures et la nature des négociations seront-elles différentes ou réparties différemment dans le discours ? L'analyse d'interactions homme-machine multimodales (grâce à la technique du Magicien d'Oz) actuellement en cours permettra de répondre à ces questions et d'élaborer un modèle de dialogue multimodal .

Bibliographie

- AUS62 J.L. Austin (1962). *How to do things with words*. Oxford University Press, UK.
- MIN75 M. Minsky (1975). *A Framework for Representing Knowledge*. dans *The Psychology of Computer Vision*, P.H. Winston, ed., McGraw-Hill, New York, N.Y.
- MON91 C. Montarnal (1991). *Etude expérimentale de l'activité de linéarisation dans les textes descriptifs de configurations spatiales*. Notes et Rapports de Recherche du CRISS no. 23, Université Mendès France, Grenoble.
- SEA69 J.R. Searle (1969). *Speech Acts*. Cambridge University Press, UK.
- SEA83 J.R. Searle (1983). *Intentionality*. Cambridge University Press, UK.
- SUC87 L.A. Suchman (1987). *Plans and Situated Action* . Cambridge University Press, Cambridge.
- WEI76 J. Weizenbaum (1976). *Computer Power and Human Reason* . Freeman.