Eléments pour la conception d'énoncés multimodaux en Dialogue Homme Machine : pourquoi l'unité d'analyse psychologique est l'Action et non l'Information

Dominique Fréard

Université Rennes 2 – Haute Bretagne

Ecole Doctorale : *« Sciences Humaines et Sociales »*

Laboratoire de Psychologie Expérimentale (LPE)

Centre de Recherche en Psychologie Cognition et Communication (CRPCC)

THESE

Présentée en vue de l'obtention du grade de Docteur en Psychologie de l'Université Rennes 2

Spécialité : *Psychologie, Cognition et Communication*

Eléments pour la conception d'énoncés multimodaux en *Dialogue Homme Machine*

Pourquoi l'unité d'analyse psychologique est l'*Action* et non l'*Information*

Dominique Fréard

2 juillet 2009

Jury

André Tricot	Professeur, Université Toulouse II	(Rapporteur)
Pierre Falzon	Professeur, CNAM	(Rapporteur)
Eric Jamet	Professeur, Université Rennes 2	(Directeur)
Philippe Bretier	Docteur en Informatique, France Télécom R&D	(Examinateur)
Ludovic Le Bigot	Maître de Conférences, Université de Poitiers	(Examinateur)

Thèse préparée au laboratoire EASY de France Télécom R&D, 2 avenue Pierre Marzin, 22307 Lannion Cedex

et au laboratoire LPE du CRPCC, place du Recteur Henri Le Moal, 35000 RENNES.

« L'homme regarde comme accident, et n'éprouve que par accident la manifestation et l'évidence des lois qui le régissent, qui le font, le défont, le conservent, l'altèrent, l'animent, et l'ignorent. Il ne sent battre son cœur que par moments critiques.

S'il tombe, il se rencontre lui-même. Il se heurte. S'il peut rêver (et même penser) à voler, à ne pas mourir, à… etc., c'est que les lois sont étrangères à sa pensée. Elles n'y sont que superficie : accident aperçu par accident. »

Paul Valéry

« Il faut se rendre compte que " vrai " et " faux ", tout comme " libre " et " non libre ", ne recouvrent absolument pas des notions simples ; mais simplement une dimension générale où ils représentent ce qu'il est juste et convenable de dire – par opposition à ce qu'il serait mal venu de dire – en ces circonstances, à cet auditoire, dans ce dessein et cette intention. »

John L. Austin

« Je vais donc affirmer que la vie sociale est une scène, non pas en une grande proclamation littéraire, mais de façon simplement technique : à savoir que, profondément incorporé à la nature de la parole, on retrouve les nécessités fondamentales de la théâtralité. »

Erving Goffman

Remerciements

Je remercie les membres du jury, André Tricot, Pierre Falzon, Philippe Bretier et Ludovic Le Bigot, d'avoir accepté de lire ma prose et de participer à ma soutenance de thèse. J'en suis très honoré.

Je remercie mes directeurs, Eric Jamet, Gérard Poulain ainsi que Laurence Perron qui m'ont accompagné au cours de ce travail. Merci à Eric de m'avoir offert cette possibilité. Merci à Gérard pour toute son humanité, dans le travail et en dehors. Merci à Laurence d'avoir repris le flambeau après qu'on nous ait privés de la présence officielle de Gérard. (Actionnaire, quand tu nous tiens…)

Je tiens également à remercier Valérie Botherel, qui n'a pas encadré officiellement ce travail, mais qui a fréquemment accepté d'interagir avec moi de façon toujours attentionnée, critique et constructive.

Je remercie tous ceux qui m'ont supporté pendant la gestation. Ceux déjà cités, ainsi que tous les amis, doctorants, stagiaires et autres colocataires. Merci notamment à tous ceux qui aiment les grandes discussions philosophico-théoriques. Vous m'avez tous beaucoup appris et j'ai adoré ça. Je pense à Sylvie avec qui les liens interdisciplinaires entre nos travaux sont vite devenus des boulevards ; et puis à tous ceux dont la passion transpire dans un œil qui pétille, notamment Joseph, Thomas, Ghislain, mais tous les autres aussi. Merci pour tous ces moments là et pour d'autres à Erwan, Fred, Lisa, Romain, Liv, Pierre, Laurent, Kris, Thomas, et aux autres lannionais, à Amaël et Naïla, Steph, Nadia, Cécile et Seb, et aux autres rennais.

Merci tout spécialement à Lisa qui a relu en détail une grande partie du manuscrit et qui a sans doute souffert sans trop vouloir me le dire. Merci à Ludovic pour une grande continuité de suivi qui m'a permis de m'orienter. Ça compte beaucoup. Merci à Liv qui m'a souvent remis les pieds sur terre et qui s'est inquiétée pour moi. Ça aussi, c'est très important. Beaucoup d'autres ont relu des petits bouts ou discuté des idées. C'était toujours utile ! A tous, désolé de tenir autant à mes idées…

Un merci tout particulier à Isabelle – l'aînée de mes sœurs, également institutrice rigoureuse – dont quelques commentaires dans le premier chapitre ont eu une influence sur tout le document.

Et puis merci Papa, merci Maman, pour leur soutien inconditionnel et pour tout le reste.

RESUME

Cette thèse s'inscrit dans le domaine du *Dialogue Homme Machine* (DHM). Elle vise l'amélioration des capacités de communication des systèmes de DHM. Dans ce but, les travaux présentés portent sur l'analyse des actes du système et sur l'évaluation de leurs effets dans un contexte d'interaction dialogique. Ils permettent de proposer des principes de conception susceptibles d'enrichir le comportement de ces systèmes.

La problématique de la conception des *actes du système* (nommés *« énoncés »* dans la thèse) est une question appliquée qui vise l'amélioration des capacités de communication des systèmes grand public, à vocation commerciale. Mais cette question suppose de disposer d'outils conceptuels propices à l'analyse et renvoie à la problématique du fonctionnement cognitif de l'*individu humain* (*« l'utilisateur »*) dans le contexte de la communication dialogique. La partie théorique permet de poser cette problématique appliquée dans un contexte interdisciplinaire (entre *linguistique*, *ingénierie* et *psychologie*). Elle présente et critique différents points de vue sur cette question pour en dégager les apports et les limites. Cette présentation permet d'opposer deux points de vue dans la partie expérimentale : (1) l'*approche pragmatique*, qui analyse les *actes des partenaires*, selon un point de vue sociocognitif, et (2) l'*approche cognitive*, qui analyse les *processus de traitement de l'information*, centrée sur le niveau individuel.

Cinq expériences, basées sur le protocole du *Magicien d'Oz*, ont été réalisées. Dans les deux premières, le système communiquait uniquement avec des *énoncés auditifs*, selon un mode vocal. Ces deux expériences ont permis de mettre en évidence l'utilité et les effets des *messages d'aide* (*aides procédurales*) et de la *syntaxe* dans les énoncés auditifs. Dans les trois expériences suivantes, le système utilisait des *énoncés audio-visuels*, sur un mode multimodal. Une *catégorisation des types d'information à présenter* a été introduite. Une *règle d'attribution* des modes de présentation (*auditif* ou *visuel*) aux différents types d'information (*'écho'*, *'réponse'*, *'relance'*) a été proposée pour concevoir des *'stratégies de présentation'* innovantes. Ces trois expériences ont permis de démontrer l'intérêt des principes d'analyse utilisés pour la *conception des stratégies de présentation*. Elles ont permis d'identifier l'importance de la *relation type-mode* pour prédire les effets produits par les actes dans l'interaction. Globalement, les résultats obtenus permettent de valider l'*approche pragmatique* contre l'*approche cognitive*. La discussion permet d'aborder les implications de ces résultats, tant sous l'angle de la conception des énoncés des systèmes de DHM que des conséquences théoriques qui peuvent être tirées de ces résultats.

Mots-clés : Dialogue Homme Machine ; Multimodalité ; Conception d'énoncés ; Charge cognitive ; Action

TABLE DES MATIERES

Contact : dominique.freard@gmail.com

En ce qui concerne l'utilisation des systèmes de dialogue,

cette thèse est une complainte d'un utilisateur malheureux.

(au sens austinien)

Contact : dominique.freard@gmail.com

INTRODUCTION

La perspective de cette thèse est l'amélioration des capacités des *systèmes de dialogue* à s'exprimer utilement auprès de leurs utilisateurs. L'objectif est de comprendre quelles connaissances, notamment en psychologie, sous-tendent la manière dont on fait actuellement parler ces *systèmes artificiels*. En effet, les développements techniques sont basés sur les connaissances en sciences-humaines, lesquelles y trouvent une raison d'être et des limites. La démarche de la thèse vise à étudier ces connaissances, à les questionner et, peut-être, à proposer des pistes d'amélioration pour la conception.

Commençons par un exemple. Si, le 10 juin 2008 à 11h, vous aviez composé le 36 35[1] sur votre téléphone (en France), voici la transcription du message que vous auriez entendu :

> *« (Jingle sonore : Ta Ta Tam Tam) … Bonjour et bienvenue à la SNCF.*
>
> <pause, 2 secondes>
>
> *Cet appel sera facturé 34 centimes d'Euros par minute.*
>
> *Suite à un mouvement social national, nous vous informons, gratuitement, sur l'état du trafic au numéro vert suivant : 08 05 90 36 35.*
>
> *Je suis à votre écoute. Précisez-moi le nom du service qui vous intéresse ; Sinon, laissez-vous guider.*
>
> <pause, 2 secondes>
>
> *Je vais maintenant vous suggérer différents services. Merci de me dire ce que vous souhaitez.*
>
> *Le SERVICE BILLET, pour ajouter ou annuler un billet. Le PROGRAMME FIDELITE, pour bénéficier de vos avantages. Je vous propose aussi de consulter l'état du TRAFIC, les HORAIRES, ou nos SERVICES PRATIQUES pour faciliter votre voyage.*
>
> <pause, 5 secondes>
>
> *Vous pouvez aussi dire GUIDE pour consulter le guide d'utilisation des services. »*
>
> *(Total : 59 secondes)*

Ce message correspond bien à l'état actuel du savoir-faire concernant la conception des *systèmes de dialogue* (services téléphoniques automatiques). Il apporte les informations nécessaires pour différents types de clients qui connaissent plus ou moins bien les services de l'entreprise. Il guide ces utilisateurs pour tenter de prendre en compte leurs besoins et il

[1] Cet exemple a été choisit du fait de la notoriété de ce service en France. Cela illustre l'aspect quotidien de la problématique. Ce service peut recevoir jusqu'à 10 000 appels à l'heure.

est de plus en plus directif de façon à aider les novices à surmonter d'éventuelles difficultés d'utilisation. Il permet ainsi à tous les utilisateurs de construire une représentation du système suffisante pour son utilisation et il guide de plus en plus profondément les utilisateurs qui connaîtraient moins bien le service, la SNCF ou les systèmes vocaux en général.

Différentes informations sont fournies à l'utilisateur du service sous la forme de neuf phrases successives qui forment une structure linéaire – la seule possible en mode vocal –. Cette structure rectiligne des *« messages »* (ou *« prompts »*, ou encore *« énoncés »*) du système donne lieu à une monotonie (bien connue des utilisateurs de ces services) qui engendre une certaine défiance des utilisateurs (ou *« clients »*) potentiels et qui freine l'adoption de ces technologies.

En termes de recherche, les pistes envisagées pour l'amélioration portent sur le contenu des énoncés (vocabulaire, syntaxe, indications apportées à l'utilisateur) et sur leur forme (prosodie de la synthèse vocale ou de la voix enregistrée, multimodalité). Le problème qui se pose est celui de la *'complexité'* des traitements imposés à l'utilisateur, que l'on renvoie fréquemment à la quantité d'information(s) à traiter, à la *'métaphore de la pénétration des informations'* et au problème de la *'charge cognitive'*. Ce point de vue, très répandu, est focalisé sur les processus de mémoire et d'apprentissage, vus comme une *« télécopie dans l'esprit »* (expression empruntée à Baker[1], 2004) des informations présentées par le *système informatique*. Ce point de vue est critiqué depuis quelques années (*e.g.* Baker, 2004 ; Bard & al., 2007 ; Gerjets et Scheiter, 2003 ; Kirsch, 2000 ; Theureau, 2002 ; Tricot & Chanquoy, 1996). La thèse s'attache à démontrer qu'il est insuffisant pour analyser les effets des énoncés dans le contexte du dialogue homme machine. L'*analyse pragmatique* des énoncés est proposée comme l'alternative qui permet de questionner la technique et de proposer des solutions susceptibles de rendre l'utilisation plus acceptable. Cette analyse consiste à identifier les actes et leurs effets dans le contexte de la communication pour établir les relations de causalité internes à ce processus. Selon ce point de vue, l'unité d'analyse n'est pas l'*information*, mais l'*acte*.

Les questions posées dans la thèse portent sur la réduction des contraintes liées au traitement des énoncés. Elles portent notamment sur la nécessité de présenter certaines informations, sur leur utilité et sur leur forme. L'objectif est d'étudier la diversité des effets des énoncés du point de vue *sociocognitif*, de façon à préciser comment faire produire aux systèmes de dialogue des énoncés adaptés aux besoins des utilisateurs. La thèse indique qu'il est possible d'accroître les compétences pragmatiques des systèmes dialogiques. Cette idée est exprimée ainsi par Karsenty (2003, p. 109) : « *Une vision enrichie de la coopération dans le dialogue homme-machine s'avère (…) nécessaire, dans laquelle le système n'adopterait pas seulement les buts informationnels de l'utilisateur mais aussi ses besoins*

[1] Précisons, même si cela peut sembler évident, que cet auteur ne défend pas ce point de vue. Il utilise cette expression pour en montrer la limite.

cognitifs liés à l'activité de dialogue. » Ainsi, la connaissance des *'besoins cognitifs liés à l'activité'* doit permettre d'envisager une communication plus riche.

L'ambition de la thèse est de contribuer à l'*analyse des énoncés des systèmes de dialogue homme machine* pour tenter d'en déduire des règles d'analyse utilisables pour la conception. De telles règles peuvent être utilisées par des *équipes de conception*, qui travaillent à l'échelle d'un projet, ou implémentées dans des *algorithmes de gestion dynamique des énoncés*, qui travaillent en temps réel dans l'interaction. Cette analyse relève d'une approche multidisciplinaire qui implique l'*ingénierie*, la *linguistique* et la *psychologie*. Cela implique que la référence à certains travaux externes à la psychologie est nécessaire pour présenter les principes de fonctionnement des systèmes, la démarche de conception et les notions théoriques qui permettent l'analyse.

La contribution des travaux de la thèse porte sur la prise en compte de la structure complète des énoncés du système, dans un cadre expérimental, pour étudier les possibilités de présenter l'information sur un mode multimodal. Les expériences proposées visent à définir des *'stratégies de présentation multimodale'* (*Cf.* Partie expérimentale) utilisables dans le contexte du dialogue homme machine. La méthodologie choisie a pour but d'obtenir un diagnostic en laboratoire. Elle ne porte pas sur l'usage réel dans la mesure où aucun service réel ne propose aujourd'hui ces modes de communication. L'utilisation de ces stratégies dans un contexte d'usage réel suppose des phases préalables de généralisation qui permettraient de relativiser les effets obtenus en laboratoire. La thèse se limite à l'étude des principes d'organisation des énoncés et aux règles d'analyse et de conception qui peuvent y être appliquées. Les questions se rapportent aux interactions et à la performance du couple homme machine. La performance est évaluée grâce à l'observation et à l'interprétation des échanges dans ce couple et grâce aux évaluations subjectives des participants et à la mémorisation des contenus. L'objectif est de proposer un cadre d'analyse des énoncés multimodaux, incluant leur conception et leur évaluation, et de l'appliquer à l'étude de quelques stratégies de présentation dans le but de démontrer sa pertinence pour le développement des *compétences pragmatiques* des systèmes de dialogue. Plus simplement, il s'agit de définir des *patterns comportementaux* utilisables par des systèmes multimodaux lorsqu'ils s'adressent aux utilisateurs.

La communication entre humains est le modèle de référence pour la conception des situations de communication homme-machine. Ici, plutôt que de chercher à reproduire ce modèle, l'idée est de s'en émanciper et de chercher des moyens qui pourraient permettre à un système automatique *'d'assumer son corps de système artificiel'* pour en tirer partie dans la communication. L'idée sous-jacente est que comme ce corps est différent, les avantages et les inconvénients qui en résultent peuvent être différents. Ces mécanismes doivent, ou peuvent, être étudiés pour eux-mêmes.

PARTIE THEORIQUE

Une cartographie des environnements interactifs dédiés à l'apprentissage, proposée par Baker (2003), distingue quatre situations dans lesquelles l'ordinateur joue un rôle différent dans la communication :

(1) *Dialogue avec les technologies* : L'ordinateur est un '*participant*' à part entière. Il agit comme un agent autonome, servant l'utilisateur avec lequel il dialogue ;

(2) *Dialogue autour des technologies* : L'ordinateur est un '*support*'. Il agit comme un objet dont peuvent se servir, et duquel peuvent parler, les différentes personnes en coprésence physique ;

(3) *Dialogue au travers des technologies* : L'ordinateur est un '*médiateur*'. Il agit comme un pont reliant des rives distantes ;

(4) *Dialogue entre les ordinateurs* : L'ordinateur est un '*participant*' autonome. Ici, l'emploi du terme « dialogue » est abusif.

Les travaux présentés dans la partie expérimentale, et plus généralement le DHM, se rapportent à la première catégorie. Dans cette situation, l'ordinateur est l'un des partenaires en communication, ce qui suppose qu'il est *autonome*. Cela implique, pour cet acteur, de disposer de capacités suffisantes pour *comprendre correctement* son partenaire, *prendre les bonnes décisions* et *agir en conséquence*. Il doit réaliser une boucle d'action complète. En ceci, la problématique associée à cette situation est la plus « complète » car elle implique l'impossibilité à se reposer, en temps réel[1], sur les décisions et les capacités d'adaptation d'un humain. Le système est autonome et doit disposer des connaissances nécessaires pour faire les bons choix stratégiques. Ainsi, la conception d'un système automatique suppose de formaliser ces connaissances de façon aussi complète que possible.

La partie théorique a pour but de circonscrire la problématique de la conception des énoncés des systèmes de *Dialogue Homme Machine* (DHM). Elle est organisée autour de la question de la conception des énoncés, mais celle-ci n'est abordée spécifiquement que dans le chapitre 3. Les autres chapitres théoriques (chapitres 1, 2 et 4) permettent de définir les méthodes d'analyse, le contexte technique, les principes de développement et les notions théoriques nécessaires dans le cadre du DHM. Cette présentation large est faite pour clarifier

[1] Des informaticiens diraient au « *run time* », *i.e.* pendant l'exécution du programme.

l'approche analytique du problème posé. Elle permet de relativiser les points de vue les uns par rapport aux autres. Elle permet également d'aborder des points méthodologiques essentiels. Tous ces éléments sont pris en compte par la suite (chapitre 5) pour proposer une synthèse théorique et une analyse des tâches étudiées et pour en dégager la problématique de la thèse.

Les quatre chapitres qui composent la partie théorique sont les suivants :

- Le chapitre 1 porte sur *l'analyse des énoncés* produit par les partenaires dans le processus de communication. Il permet de définir ce qu'est l'énoncé dans le but de comprendre de quoi il se compose, pourquoi on l'emploie et quels en sont les effets. Ce chapitre présente d'abord la *théorie des actes de langage* (Austin, 1962), l'*analyse conversationnelle* et la *théorie de la collaboration* (Clark et ses collaborateurs). Il aborde ensuite la question de la *hiérarchie des processus* nécessaires à la production et à la compréhension des énoncés, puis la question de la *pertinence*.
- Le chapitre 2 est un état de l'art de la *conception des systèmes de DHM*. Il présente les bases technologiques qui permettent à des systèmes informatiques d'exprimer des énoncés auprès d'utilisateurs. Ce chapitre permet d'abord de donner un aperçu du fonctionnement des systèmes de dialogue. Il aborde ensuite les questions liées à la conception et s'oriente sur la conception des énoncés du système. La question de la multimodalité est abordée dans la troisième partie.
- Le chapitre 3 permet d'aborder les travaux empiriques sur *la conception des énoncés* des systèmes interactifs. Il permet de mieux situer la problématique de la conception des énoncés dans une perspective psychologique. La première partie est consacrée à l'étude des énoncés en DHM. La seconde partie porte sur le domaine de la psychologie des apprentissages à partir de systèmes multimédias. Enfin, une courte partie sur la redondance entre modalités permet d'introduire l'*effet de préemption* (Helleberg & Wickens, 2003).
- Au fil des chapitres 2 et 3, *la notion de charge cognitive*[1] de l'utilisateur est récurrente. Le chapitre 4 revient sur les fondements de cette notion et sur les développements auxquels elle donne lieu. Ce chapitre présente les problèmes liés à la définition et à la mesure de la charge cognitive. Il s'oriente ensuite sur le rôle causal attribué à la notion, au nom duquel elle est utilisée pour expliquer les performances des individus. Une approche critique de ce problème conduit vers la problématique de la thèse.

[1] L'expression *'charge cognitive'* est utilisée de façon préférentielle dans ce document. Dans la littérature, on trouve également les expressions *'charge de travail'* et *'charge de travail mentale'* selon les préférences des auteurs et selon le domaine. Dans la mesure du possible, les termes utilisés dans la partie théorique correspondent à ceux employés par les auteurs cités.

Chapitre 1 L'analyse pragmatique des énoncés

L'analyse pragmatique des *énoncés* linguistiques a été proposée initialement par Austin (1962). Elle est dite *pragmatique* parce qu'elle porte sur les actes produits dans la communication. Elle permet de mettre en évidence la structure et les effets des *énoncés*, ainsi que les règles qui jouent dans leur interprétation.

A partir de l'analyse d'Austin, divers travaux ont permis d'élargir le point de vue à d'autres aspects de la communication, plus globaux que l'énoncé, tels que *'la référence'*, ou *'l'interaction'* ; et de s'intéresser aux connaissances mises en place par les interlocuteurs, notamment avec la notion de *'terrain commun'* et l'étude du *'processus de grounding'.* – Ces notions sont abordées dans le cœur de ce chapitre. – Avec ces avancées, l'analyse pragmatique est aujourd'hui la base de travaux sur des notions telles que, par exemple, *l'argumentation* ou *l'apprentissage* pour lesquels elle offre une perspective large (voir, par exemple, Baker, 2004, qui propose de faire converger ces deux domaines).

L'objectif de ce chapitre est de présenter la richesse de l'analyse permise grâce à ces divers développements. Il est présenté dans une perspective historique qui a pour objectif de montrer comment la discipline s'est établie à partir de la notion d'*énoncé* avant de s'en émanciper ; et quel retour sur la notion est possible aujourd'hui grâce à ces élargissements de point de vue.

1.1 Origines

Le terme « pragmatique » renvoie aux actes produits dans le monde réel et à leurs effets. En tant que discipline, la pragmatique peut donc renvoyer à plusieurs champs scientifiques ayant trait à l'action, à la communication et à leurs effets : philosophie, linguistique, psychologie et sociologie, notamment.

Ses origines sont lointaines puisqu'elles datent des premières réflexions sur la rhétorique que l'on trouve chez les grecs anciens. Notamment, Aristote classait les discours en trois genres : *« judiciaire »* pour le jugement des actes passés, *« épidictique »* pour le discours sur les faits présents, et *« délibératif »* pour un discours engageant des décisions pour l'avenir. Il cartographiait les arguments rhétoriques et proposait, en complément, une méthode dialectique pour établir les principes de l'argumentation (pour une synthèse, voir Schopenhauer, 1831). La pragmatique moderne opère un retour à ces préoccupations liées au discours oral et à l'argumentation. C'est pourquoi la notion d'*énoncé*, ou d'*énonciation*, y fait l'objet d'un intérêt particulier.

1.2 Théorie des actes de langage

La pragmatique s'est développée sur la base de la *théorie des actes de langage*, qui analyse la *parole* en tant qu'*action* dirigée vers un auditoire dans un contexte particulier. Il ne s'agit pas d'une théorie linguistique à proprement parler, car la linguistique analyse les structures grammaticales (voir, par exemple, Caron, 1989), mais, comme son nom l'indique, d'une théorie des actions réalisées grâce à la parole, *i.e.* elle analyse *les comportements d'utilisation de la parole*. En ceci, elle est très éclairante dans le domaine psychologique.

Cette théorie a été constituée par Austin (1962), puis Searle (1969), du point de vue analytique, ainsi que par Grice (1957; 1969; 1975) qui a initié l'analyse des aspects intentionnels des actes. Ces auteurs ont poursuivi une réflexion en philosophie analytique débutée par Frege (1848-1925) et Russel (1872-1970) dont les objectifs étaient de proposer une méthode de formalisation des concepts basée sur la logique et une théorie générale de la rationalité. Il s'agissait de représenter la langue sous forme de formules logiques[1] pour mettre en évidence la consistance interne du système linguistique, dans une théorie générale. – Ces réflexions ont fait naître la perspective du traitement automatique des langues dans des modèles formels (*Cf.* chapitre 2). – Sur cette base, Wittgenstein (1922) montre que tout énoncé repose sur un système formel a priori, et est, en conséquence, tautologique. Autrement dit, la valeur de vérité d'un énoncé doit être recherchée uniquement dans le système formel qui permet de l'exprimer : le langage. Il n'est pas vrai *en soi*. Il est vrai relativement à ce code. Pour cette raison, la valeur de vérité de l'énoncé ne permet pas de comprendre les rapports qu'il entretient avec la réalité du monde et *sa signification ne peut être déduite directement du signal*. Cette démonstration a eu pour conséquence de convaincre plusieurs auteurs que, pour développer une *science de la signification*, une analyse approfondie des emplois du langage était nécessaire.

Par ailleurs, cette analyse connaissait des avancées dans le domaine de la linguistique avec les travaux de Peirce (1839-1914). Pour lui, tout élément ou évènement, de l'individu ou de l'environnement, peut être un signe à partir du moment où un interprète le prend en considération. On peut, par exemple, voir des signes dans les nuages. De ce fait, plutôt qu'au signe en lui-même, il s'intéresse à la *mise en signe*. Il dépasse le rapport binaire entre '*signifiant*' et '*signifié*' (*i.e.* le mot et le sens), classique en linguistique structuraliste (Saussure, 1913). Pour lui, la mise en signe est une opération ternaire entre un signe matériel (qu'il nomme *representamen*), un objet (le *référent*) et un *interprétant* (l'individu responsable de l'association entre le *representamen* et le *référent*). Dans cette opération, l'individu vient prendre sa place entre la *langue* et le *monde*. Ainsi, à partir de ces travaux, il devenait important d'expliquer le rôle actif de l'individu *dans l'emploi de la langue* et *pour l'action dans le monde*. C'est ce que proposera Austin.

[1] L'ensemble des formules logiques qui sont nécessaires à la représentation d'une langue forment un '*système formel*'.

1.2.1 Austin et l'énonciation

C'est en comparant divers énoncés linguistiques qu'Austin a établi les fondements de la *théorie des actes de langage*. Ces travaux ont été exposés dans une série de conférences (université de Harvard, 1955) dont les notes ont été éditées après sa mort (Austin, 1962). Ce texte est essentiel pour l'analyse des fonctions remplies par un énoncé dans un échange à visée communicative.

A Les énoncés performatifs

Jusqu'à Austin, toute *énonciation* était vue comme une *affirmation*, ce qui supposait que son sens était toujours vérifiable. L'affirmation était *vraie* ou *fausse*. Les cas d'exception avaient été traités par Kant, qui considérait les *'fausses affirmations'* comme des phrases de structure grammaticale correcte mais dont le contenu propositionnel aurait été un non-sens. Pour Austin, cette dichotomie entre *'fausses affirmations'* et *'affirmations classiques'* doit être rejetée car elle est trop simple. Il désigne la croyance qu'une affirmation devrait toujours être vérifiable sous le nom « *d'illusion descriptive* » et il la qualifie de *« dogmatique »*.

Austin distingue la notion d'*énonciation performative* (*« performative utterance »*) de celle d'*affirmation* (« *statement »*, ou *« sentence »*). Les *énonciations performatives* sont des expressions utilisant des verbes tels que : parier, remercier, se marier, baptiser, etc. Il fait remarquer que ces verbes ne sont pas destinés à produire des affirmations et qu'ils n'ont pas de valeur de vérité. Pour les interpréter, le contexte de l'énonciation joue un rôle important. Austin parle de *« circonstances appropriées »*. Par exemple, si votre petite sœur déclare que vous êtes marié(e) à sa poupée, vous pouvez considérer que vous n'êtes pas réellement engagé(e) dans ce mariage. Pourtant, ce qu'elle dit en célébrant la cérémonie ne peut pas être qualifié de faux. On peut simplement dire que l'effet visé par son affirmation ne s'est pas produit. Il s'agit d'un *échec* (*« infelicity »*).

B Notion de réussite/échec de l'énonciation

La notion d'échec permet d'analyser les rapports entre l'*énoncé* et les *circonstances de l'énonciation*. Austin cartographie les *« choses qui peuvent se mal présenter et fonctionner mal »* pour une énonciation performative (Tableau 1-1). Cette classification montre quelles conditions doivent être remplies pour que l'acte puisse être considéré comme *heureux*. Si toutes les conditions sont remplies, l'acte réussit et il est dit *'heureux'* ; si elles ne sont pas remplies, il échoue et il est dit *'malheureux'*. On parle de '*conditions de félicité' de l'acte*.

Le Tableau 1-1 présente le classement des réussites/échecs proposé par Austin. La première colonne désigne le mode de *'classement'* qu'il utilise. La deuxième colonne nomme les différents *'types d'échecs*' de ce classement, conformément aux termes utilisés par Austin. La troisième colonne décrit les *'conditions de félicité'* (de *réussite)* qui correspondent, par opposition, à chaque *type d'échec*. Chacune de ces conditions doit être remplie pour qu'un acte puisse être considéré comme *heureux*.

Tableau 1-1 : Les différents types d'échecs d'un acte

AB	**INSUCCES**, Acte prétendu mais vide	
A	***Appels indus**, Acte interdit*	
A.1	*« Absence de convention »*[1]	*« Il doit exister une procédure, reconnue par convention, dotée par convention d'un certain effet, et comprenant l'énoncé de certains mots par de certaines personnes dans de certaines circonstances. »*
A.2	Emplois indus	*« Il faut que, dans chaque cas, les personnes et circonstances particulières soient celles qui conviennent pour qu'on puisse invoquer la procédure en question. »*
B	***Exécution ratée**, Acte vicié*	
B.1	Défectuosités	*« La procédure doit être exécutée correctement par tous les participants. »*
B.2	Accrocs	*« La procédure doit être exécutée intégralement par tous les participants. »*
Γ	**ABUS**, Acte purement verbal, mais creux	
Γ.1	Insincérité	*« Lorsque la procédure suppose chez ceux qui recourent à elle certaines pensées ou certains sentiments, lorsqu'elle doit provoquer par la suite un certain comportement de la part de l'un ou l'autre des participants, il faut que la personne qui prend part à la procédure ait, en fait, ces pensées ou sentiment, et que les participants aient l'intention d'adopter le comportement impliqué. »*
Γ.2	*« Défaut d'exécution »*[2]	*Ils doivent se comporter ainsi, en fait, par la suite.*

Deux résultats sont alors obtenus : (1) Grâce à divers exemples, Austin montre que la notion d'échec peut être appliquée à tous les types d'énonciation. Les *affirmations classiques* peuvent également connaître les différents types d'échecs. (2) Puisqu'une énonciation performative est une action, elle doit être exécutée par une personne. Cela suppose l'emploi d'un verbe à la première personne et renvoie à une convention entre les personnes sur la manière dont l'acte doit être interprété, et parfois à une séquence d'actions qui suit l'acte. Ainsi, expliquer un acte de langage revient à expliquer comment il doit être reçu, au travers du *cadre conventionnel* qui régit l'acte et de l'*effet visé* par le locuteur.

C Les trois niveaux de l'acte

Austin résume en baptisant les trois niveaux de l'acte qu'il a mis en évidence (la Figure 1-1 propose une schématisation de ces trois niveaux) : (1) La dimension *locutoire* correspond à l'acte verbal tel qu'il est vu habituellement par la linguistique. Il renvoie à un sens qui peut, ou non, être compris ; (2) La dimension *illocutoire* correspond à ce qui est effectué « *en* » disant. L'enseignant enseigne, l'élève récite, les amis échangent, etc. Cela correspond à l'intention du locuteur d'accomplir un acte au travers d'une convention ; (3) La dimension *perlocutoire*,

[1] A.1 et Γ.2 ne sont pas dénommés par Austin. Les termes *« absence de convention »* et *« défaut d'exécution »* sont proposés ici sur la base des explications de l'auteur.

[2] Cf. note précédente.

correspond à ce qui est effectué *« par »* le fait de dire ; par exemple, inviter l'interlocuteur à répondre, ou encore, tenter de le convaincre. Sont mentionnés les effets sur les *sentiments*, les *pensées* et les *actes*. Mais l'effet perlocutoire ne porte que sur la compréhension qui résulte de l'acte, indépendamment des conséquences. L'auteur distingue *« objectif perlocutoire »* (la compréhension) et *« suite perlocutoire »* (la réponse obtenue).

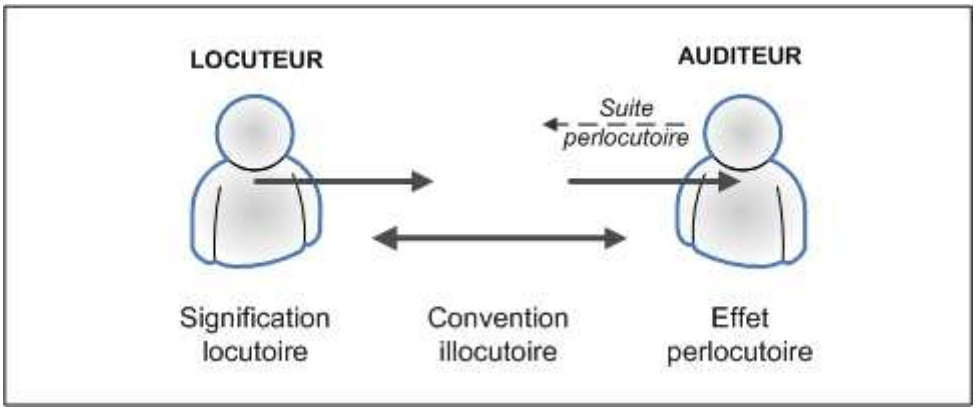


Figure 1-1 : Les trois niveaux de l'acte de langage

Austin envisage une « théorie des valeurs illocutoires » pour synthétiser les différentes *fonctions* réalisables par le biais des actes de langage. Il revendique également une théorie générale de l'action et propose une définition : *« L'acte est généralement tenu pour un évènement physique précis, effectué par nous, et distinct à la fois des conventions et des conséquences. »* On notera que cette définition peut inclure les actes cognitifs qui, même s'ils ne sont pas perceptibles, sont bien des évènements physiques dans la théorie cognitive.

D Incidences

En conclusion, tout acte de communication est composé des trois dimensions : locutoire, illocutoire et perlocutoire. Tout acte peut être décrit à la fois en termes de 'réussite *vs* échec' et de 'vérité *vs* fausseté'. Ces dimensions générales s'expriment à des degrés divers dans les différents énoncés de la langue. Austin note alors que la distinction initiale entre *performatif* et *constatif* (autre nom pour les affirmations classiques) était une première synthèse trop simple, liée à une analyse de surface. Il propose un classement des valeurs illocutoires des actes correspondant aux verbes trouvés dans le dictionnaire : (1) *les verdictifs* : un verdict est rendu (jugement) ; (2) *les exercitifs* : un pouvoir est exercé ; (3) *les promissifs :* une promesse est faite ; (4) *les comportatifs* : un comportement social est adopté ; (5) *les expositifs* : un argumentaire est exposé.

Finalement, trois enseignements concluent l'ouvrage :

(1) *Tout acte doit toujours être considéré dans son ensemble* ;

(2) *L'affirmation est un type d'acte parmi d'autres*, sans position privilégiée ;

(3) *La théorie de la signification ne peut pas ne pas tenir compte de ces faits*.

1.2.2 Points de vue complémentaires à celui d'Austin

A Searle et les actes illocutoires

A la suite des premiers travaux d'Austin, Searle (1969) a contribué à étendre cette théorie dans une perspective de rationalisation des règles du discours et d'une formalisation en un modèle logique. Cette formalisation est développée notamment par Searle et Vanderveken (1985), puis Vanderveken (1988). Leur but est de s'extraire des critères de la langue pour décrire objectivement les *actes illocutoires*[1]. Ils en viennent alors à proposer des *descripteurs* de la *force illocutoire*. D'une douzaine de descripteurs dans la première version (Searle, 1969), les travaux ultérieurs (Searle & Vanderveken, 1985) ont permis de réduire cette liste à six dimensions essentielles (*Cf.* Tableau 1-2).

Tableau 1-2 : Les composantes de la force illocutoire

(1) Le but de l'acte	*correspond à ce qui est réalisé*
(2) Le degré de puissance	*dépend de l'engagement du locuteur qui peut, par exemple, demander ou exiger*
(3) Le mode d'accomplissement[2]	*est le moyen mis au service du but, tel que le fait de laisser une option de refus lors d'une demande ou le fait de se conformer aux règles de politesse*
(4) Les conditions de contenu propositionnel	*doivent être en rapport avec l'acte. On ne demande pas de fermer une porte déjà fermée*
(5) Les conditions préparatoires	*sont les présuppositions du locuteur sur la capacité de l'interlocuteur à interpréter l'énoncé*
(6) Les conditions de sincérité	*supposent que le locuteur exprime honnêtement ses croyances, désirs et intentions*

Ces dimensions permettent de caractériser tout acte de langage, soit tout énoncé. Searle (1969) les utilise pour affiner le classement proposé par Austin (1962). Les cinq catégories qu'il propose sont : (1) *les assertifs* ; (2) *les directifs* ; (3) *les promissifs* ; (4) *les expressifs* ; (5) *les déclaratifs*. Par ailleurs, le travail de formalisation a suscité un grand intérêt en intelligence artificielle qui a permis le développement de réalisations techniques (*e.g.* Sadek, 1991; Sadek, Bretier & Panaget, 1997; Traum & Allen, 1991) qui seront mentionnées au chapitre 2.

B Les "implicatures" gricéennes et le principe de coopération

Les travaux de Grice (1957; 1975) sur l'interprétation des *énoncés* apportent un point de vue un peu différent. Grice (1957) notait l'insuffisance du modèle codique pour expliquer la communication, car l'intention intégrale d'un locuteur n'est pas codée dans l'énoncé qu'il

[1] L'expression *« acte illocutoire »* est un abus de langage. La dimension illocutoire n'est, en réalité, qu'une composante de l'acte, une partie différente du tout. Ainsi, l'expression *'acte illocutoire'* désigne l'*acte de langage* en évoquant une seule de ses parties. Il s'agit d'une métonymie.

[2] Les linguistes parlent de *« modalité »* pour désigner le mode d'accomplissement. Il s'agit bel et bien d'une modalité linguistique telle qu'elle est exprimée ici, et qui ne doit pas être confondue avec la modalité perceptive (auditive ou visuelle le plus souvent) telle qu'on la conçoit en psychologie cognitive.

produit. Il considérait nécessaire de faire intervenir un processus inférentiel parallèle. Ce processus permet de décomposer l'acte en trois intentions, auxquelles Grice (1969) fait référence sous le nom de « m-intention » (*Cf.* Tableau 1-3). A partir de cette idée, il remarque d'abord que le langage naturel ne peut pas être considéré comme totalement intelligible puisqu'il n'est souvent pas possible d'attribuer une valeur de vérité définitive aux expressions utilisées. Il introduit alors la notion d'*implication* (*« implicature »*) pour désigner l'opération d'attribution d'une signification par le destinataire à partir de l'énoncé du locuteur. Les implications permettent de démarrer de *prémisses* pour en déduire des *conclusions*. Pour Grice, elles permettent au destinataire du message de *construire le sens de l'énoncé*.

Tableau 1-3 : Signification d'un énoncé (m-intention)

(1) Le locuteur a l'intention que, du fait de son énoncé, l'auditeur produise une réponse.
(2) Le locuteur a l'intention que l'auditeur pense, au moins en partie du fait de son énoncé, que le locuteur attend une réponse de l'auditeur.
(3) Le locuteur a l'intention que la réalisation de son intention (2) soit au moins en partie la cause de la réalisation de son intention (1).

Pour Grice (1975), les implications conversationnelles sont des propriétés du discours qui utilisent un principe général des échanges verbaux : le *principe de coopération*. Ce principe est le suivant :

> *« A chaque étape, certains choix conversationnels possibles seront exclus comme conversationnellement inadéquats. Nous pouvons donc formuler approximativement un principe général que les participants seront censés respecter (toutes choses égales par ailleurs) : que votre contribution à la conversation, au moment où elle intervient, soit conforme au but ou à la direction acceptée de l'échange verbal auquel vous participez »*

Ce principe général, qui régit les conversations humaines, est exprimé par l'auteur sous la forme d'un ensemble de maximes empruntées à la philosophie (Tableau 1-4).

Tableau 1-4 : Les maximes conversationnelles

Maximes de quantité	(1) *Etre suffisamment informatif,* (2) *Ne pas être plus informatif que nécessaire.*
Maximes de qualité	(1) *Ne pas dire ce que l'on croit faux,* (2) *Ne pas dire ce que l'on n'a pas de raison de croire*
Maxime de relation	(1) *Etre pertinent*
Maximes de manière	(1) *Eviter d'être obscur,* (2) *Eviter l'ambiguïté,* (3) *Etre bref,* (4) *Etre ordonné.*

Les maximes ont pour objet de définir la pertinence d'une énonciation sur la base des interprétations qu'elle permet. On peut noter que le non respect des maximes dans une conversation permet également des interprétations. Elles seront utilisées par la suite dans le but de construire une logique applicable au dialogue. Par exemple, Gordon et Lakoff (1975) ou Gazdar (1979) ont tenté d'utiliser les maximes comme des postulats conversationnels qui

prennent les énoncés comme prémisses et en déduisent une représentation pragmatique de l'acte. Ce type de démarche vise l'implémentation du modèle dans un système interactif. Mais, comme le remarquent Sperber et Wilson (1989), cette démarche ne s'applique qu'à des cas bien délimités pour lesquels il est possible de poser une règle, ce qui constitue une connaissance *a posteriori*. La conception de dialogues automatiques suppose de ce fait une spécification complète de dialogues dédiés à des tâches spécifiques. Par ailleurs, la maxime de relation est redondante avec l'objectif des maximes qui porte lui-même sur la pertinence des énoncés.

On voit ainsi que, bien qu'elles aient des vertus empiriques, les maximes conversationnelles ne permettent pas de déduire un modèle explicatif de la communication (Sperber & Wilson, 1989). Dans la perspective du traitement automatique des langues (en compréhension et/ou en production), elles ne sont pas suffisantes pour déduire des règles qui permettraient une gestion automatique des conversations.

C Conclusion

Malgré les efforts de ces auteurs, certaines critiques peuvent être adressées à la théorie des actes de langage dans sa forme initiale. D'une part, *l'aspect méthodologique* n'est pas développé dans la mesure où cette théorie est basée sur des exemples de la langue proposés par les auteurs, sans démonstration expérimentale (Allwood, 2006; Blanchet, 1995; Sperber & Wilson, 1989). Par ailleurs, certains auteurs, émanant de la sociolinguistique, insistent sur les règles qui régissent la conversation et sur le *'rôle des interactions'* (*e.g.* Allwood, 1976; Allwood, 1977; Goffman, 1974). Pour eux, ces aspects n'étant pas suffisamment développés dans la théorie, l'énoncé est limité à sa forme linguistique. La notion d'énoncé semble donc suffisamment définie, mais pas suffisamment intégrée dans l'environnement et dans *le contexte*.

1.2.3 Analyse des interactions

La prise en compte de l'aspect dynamique des interactions est un enjeu important pour la théorie. Schegloff (2000; 2006) présente les interactions comme l'*expérience fondamentale de la socialité* (*« fundamental embodiment of sociality »*). Il les compare à une infrastructure dont la flexibilité et la robustesse supporte la macrostructure sociale. D'un point de vue méthodologique, Goffman (1974) note une tendance à ne retenir que la partie verbale pour l'analyse d'un énoncé, alors qu'il s'agit d'un ensemble *mot-geste*. Par exemple, en réponse à la question : « Avez-vous l'heure ? », dans une réponse telle que : « Oui, il est cinq heures. », le « oui » peut sembler redondant avec la réponse elle-même. Pourtant, *« on peut imaginer une scène où B, passant près de A assis dans sa voiture garée le long du trottoir, veut qu'il soit bien clair qu'il va honorer la demande que lui fait A, mais s'aperçoit que, pendant le temps nécessaire pour consulter sa montre, il s'éloigne d'un ou deux pas de la voiture et risque ainsi de donner l'impression qu'il refuse le contact. Dire « oui » constitue alors un*

moyen tout disponible de montrer que la rencontre est bien ratifiée et qu'elle sera maintenue jusqu'au terme de sa fonction. » (Goffman, 1987, page 43). Pour cet auteur, et d'autres (notamment l'école de Palo Alto, *e.g.* Bateson, 1973), les *rites d'interaction* sont au premier plan des communications et leur analyse est nécessaire pour une compréhension plus complète du processus de communication. Allwood (2007) remarque que toutes les formes d'études empiriques présentent de l'intérêt : ethnographiques, expérimentales, interviews ou questionnaires.

Cependant, les principales ressources (historiques) dans ce domaine reposent sur des bases ethno-méthodologiques, avec l'*analyse conversationnelle*. Dans cette technique, les conversations sont enregistrées, retranscrites, structurées, comparées, etc. L'analyse permet de dégager les principales structures.

A un niveau local, les unités de description sont :

- Le '*tour de parole'*, qui correspond à l'énoncé. Il débute par une *prise de parole* et se termine par une *cession de parole* ;
- Et la paire adjacente '*question – réponse'* correspondant à deux tours de parole successifs fonctionnellement liés dans l'interaction.

Au niveau global, certaines phases présentes dans toute conversation :

(1) Une *introduction* ;

(2) Un *corps* ;

(3) Une *clôture*.

L'introduction et la clôture utilisent classiquement des saluts et formules de politesses, mais peuvent prendre bien d'autres formes (Schegloff & Sacks, 1973). Le corps des conversations quant à lui, est principalement décrit en termes de rythme soutenu ou intermittent.

A Comportements de prises et cessions de parole

Le jeu de la conversation repose avant tout sur les prises d'initiative des participants. Comme le remarquent Sacks, Schegloff et Jefferson (1974) la *prise de tour* (« turn taking ») dans l'interaction est un comportement commun à de nombreuses activités, et qui peut être utilisé comme indice de l'organisation sociale (proximité, relations de dominance, etc.). Ces auteurs notent qu'une caractérisation des tours de parole peut avoir le double avantage (1) d'être un mode de description indépendant du contexte (en tant qu'unité descriptive), et (2) d'être à la fois très sensible au contexte en ce qui concerne le résultat des descriptions. Ils référencent alors un ensemble d'effets structurels régissant les tours de parole, identifiés dans des conversations enregistrées. Ces effets sont listés dans le Tableau 1-5.

Etant donné leurs observations Sacks, Schegloff et Jefferson (1974) concluent que le système de gestion des tours de parole est :

(1) Un *système de gestion local* (*« local management system »*) ;

(2) A*dministré par les individus* (*« party-administered »*) et ;

(3) C*ontrôlé interactivement* (*« interactionally controlled »*).

En effet, chaque tour est orienté vers le suivant, sans planification préalable. Le système consiste à gérer les prises de parole localement. Il est administré indépendamment par chaque participant à la conversation, du fait de ses initiatives et en fonction des déterminismes qui lui sont propres. Toute initiative est effectuée à l'aide d'un *jeu de règle* (*« rule-set »*). Chaque tour de parole offre des options qui détermineront les tours suivants de la part des autres intervenants. Par exemple, il est obligatoire de parler quand on vous donne la parole. Dans ce système *interactif,* les énoncés produits sont conçus avec une taille minimale (variable en fonction des situations et/ou des individus), et sont interruptibles ou extensibles à certains points, selon les demandes des interlocuteurs. Des *points de transition* (*« transition places »*) peuvent donc être identifiés, ce qui en fait des éléments discrets.

Tableau 1-5 : Organisation des changements de locuteur en conversation

(1) Les changements de locuteurs sont récurrents, ou au moins, occurrents.
(2) En majorité, une seule personne parle à la fois.
(3) Il arrive que plus d'une personne parle à la fois, mais ces occurrences sont brèves.
(4) Les transitions d'un tour au suivant sans interruption ni chevauchement sont courantes. Si on y inclue les transitions avec une légère interruption ou un léger chevauchement, cela compose la majorité des transitions.
(5) L'ordre des tours n'est pas fixe, mais varie.
(6) La longueur des tours n'est pas fixe, mais varie.
(7) La longueur des conversations n'est pas fixe, mais varie.
(8) Ce qui est dit par les locuteurs n'est pas fixe, spécifié à l'avance.
(9) La distribution relative des tours n'est pas fixe, spécifiée à l'avance.
(10) Le nombre d'individus peut changer.
(11) La parole peut être continue et discontinue.
(12) Des techniques d'allocation des tours sont utilisées : le locuteur peut sélectionner le suivant, un individu peut aussi s'auto-sélectionner en prenant la parole.
(13) Différentes "unités de construction de tour" sont employées. Un tour peut être de la longueur d'un mot, ou, par exemple, d'une phrase.
(14) Des mécanismes de réparations des erreurs d'allocation de tours de parole existent. Par exemple, si deux individus prennent la parole en même temps, l'un d'eux va s'arrêter pour réparer le trouble.

Ces observations montrent que la conversation est un processus particulièrement malléable dans sa structure. Pour reproduire artificiellement un tel système, il est nécessaire de l'organiser selon diverses problématiques : tâche que s'est assigné Schegloff.

B Analyse des conversations

Pour répondre aux objectifs qu'il s'est fixé, l'auteur (*e.g.* Schegloff, 2006) conserve une perspective fonctionnelle dans ses travaux plus récents. Ceux-ci décrivent la conversation sous la forme d'*organisations génériques de pratiques conversationnelles*, à la manière des

jeux de langage (Wittgenstein, 1953). Pour chaque problème générique, des solutions pratiques sont proposées. Les problèmes référencés sont les suivants :

- Le problème des ***prises et cessions de parole*** (*« turn-taking »*) : Qui parle ? Quand ? Quelles sont les conséquences sur la construction des actes ? Ce travail vient d'être présenté (*Cf.* Tableau 1-5).
- Le problème des ***séquences organisationnelles***. Ce problème est celui de la cohérence générale et du cours d'action en fonction de ce que font les individus : demander, inviter, offrir, etc. Des séquences structurées sont identifiables. La plus simple est une *paire adjacente*, telle qu'un salut, qui implique simplement le salut du locuteur et la réponse du destinataire. Des séquences plus étendues peuvent impliquer des pré-invitations ou des préannonces qui permettent de lancer des séquences d'explications. Comme les rimes, des paires peuvent être croisées ou embrassées, et des séquences parfois complexes peuvent être identifiées (jusqu'à une centaine de lignes de transcriptions, d'après Schegloff, 1990).
- Le problème des ***troubles***. Ils peuvent concerner l'audition, la compréhension, la parole, ainsi que le maintien de l'intersubjectivité et le progrès de l'activité. L'auteur note que de nombreux troubles sont mentionnés et réparés par le même locuteur dans le même tour de parole ou dans l'*espace de transition* immédiat. Cette remarque est importante car elle implique l'existence et la gestion d'un niveau d'organisation inférieur à celui du tour de parole (ou de l'énoncé).
- Le problème de la sélection des ***mots***. Le choix du vocabulaire ne relève pas uniquement de la composante sémantique. D'une part, le choix des termes fait l'objet de négociations, et d'autre part, l'emploi de certains termes peut avoir des conséquences interactionnelles directes, telles qu'emporter l'adhésion ou la réprobation, porter au rire, etc.
- Le problème de l'***organisation de la structure globale***. Certaines actions doivent avoir lieu à certains moments. Pour commencer ou clore une interaction par exemple, les salutations et les aux revoir ne sont que deux solutions parmi un ensemble de possibilités.

Ces travaux n'ont subi que peu de modifications depuis les études initiales (*e.g.* Sacks et al., 1974). Ils sont par ailleurs régulièrement cités pour la qualité de l'approche et la généralité des structures décrites. Ces structures peuvent être utilisées autant pour l'analyse que pour la conception des situations de communication « artificielles » (*'communication homme machine'* et, dans la perspective de cette thèse : *'dialogue homme machine'*).

C Le modèle hiérarchique

Le modèle hiérarchique relève de l'*analyse du discours* plutôt que de l'*analyse conversationnelle*. Cette approche s'oriente plus sur la structure dialectique qui permet l'argumentation que sur l'analyse des *interactions,* qui sont plus prégnantes dans la

conversation. Le modèle hiérarchique est inspiré de propositions en provenance de l'école de Birmingham (*e.g.* Sinclair & Coulthard, 1975) et sa version la plus achevée est proposée par Roulet et ses collaborateurs (Roulet, 1981; Roulet, Auchlin, Moeschler, Rubattel, & Schelling, 1985) de l'école genevoise. Le modèle est dit hiérarchique dans la mesure où les différents niveaux entretiennent entre eux des relations d'inclusion et de subordination. C'est en ce sens que l'approche peut être dite *structuraliste*.

Il s'agit par ailleurs d'un modèle *fonctionnel* puisque les différentes unités pertinentes à un niveau sont dotées de fonctions à ce niveau (illocutoire, interactive, etc.). Dans sa présentation du modèle Kerbrat-Orecchioni (1995) retient cinq niveaux (*Cf.* Tableau 1-6) les plus fondamentaux, et les plus proches du consensus entre auteurs. Cette description n'exclue évidemment pas les divergences entre les auteurs, tant du point de vue des unités retenues que de celui de la terminologie qui permet de les décrire.

Tableau 1-6 : Les cinq rangs du modèle hiérarchique

L'Interaction	*C'est l'unité de rang supérieur, correspondant à la conversation. Goffman parle de rencontre. C'est l'unité ultime de l'analyse.* *Elle est décrite par le schéma de participation des différents interlocuteurs, par l'unité de temps et de lieu, par un critère thématique et par l'existence de séquences qui permettent de démarquer cette conversation.*
La séquence	*C'est un bloc d'échanges à forte cohérence sémantique et/ou pragmatique. Certains parlent, par exemple, d'épisode, de phase ou de section.* *Il s'agit d'un macro-acte permettant de traiter un sous problème dans l'interaction. On trouve, par exemple, des séquences d'ouverture ou de clôture. Le corps de l'interaction peut lui-même se composer de diverses séquences. De plus, certaines séquences peuvent être enchâssées.*
L'échange	*C'est la plus petite unité dialogale, qui occupe de ce fait un statut privilégié dans la littérature. Elle correspond à la paire adjacente décrite par Sacks et Schegloff.* *Les auteurs notent qu'il existe de multiples types d'échanges et que la recherche n'en est qu'à ses débuts.*
L'intervention	*C'est l'unité monologale (émise par un seul locuteur) la plus large. Elle correspond à l'énoncé. Elle peut dépasser le tour de parole, par exemple, si l'enfant débute une intervention par une faute de grammaire, corrigée par l'adulte avant que l'enfant ne termine.* *Pour Roulet, ce niveau d'analyse mérite une attention particulière car il s'agit du point de jonction entre le monologal et le dialogal.*
L'acte de langage	*C'est l'unité minimale de la grammaire conversationnelle. Elle renvoie à l'analyse austinienne de la fonction illocutoire. Pour Roulet, la « fonction illocutoire » renvoie aux relations entre interventions, au sein de l'échange (au niveau supérieur), alors que les relations entre les actes, au sein de l'intervention (à ce niveau) seraient décrites par une « fonction interactive ».* *A ce niveau, la question fondamentale porte sur la formulation d'une théorie de l'action. Cela impose un découpage des actes, ce qui suppose une rupture de la continuité de l'expérience et de l'être. Pour faire ce découpage, la prise en compte des indices déictiques gestuels et corporels, ainsi que de la prosodie, est indispensable. Une autre question importante porte sur l'explication des actes indirects.* *Il est entendu qu'un seul acte peut réaliser des fonctions diverses.*

Du point de vue analytique, le bénéfice de ce modèle est évident. Il fait ressortir l'aspect hiérarchique des actes réalisés dans la conversation et dans l'interaction. Ainsi, il permet de reconsidérer l'analyse austinienne, qui portait seulement sur le niveau de l'énoncé (ou « intervention » ici). L'acte de langage n'est plus vu comme un acte isolé, mais comme un élément d'un système plus vaste : l'ensemble de la conversation (ou de la rencontre).

D Bénéfice de l'analyse des conversations

Finalement, ces travaux présentent l'avantage de se fonder sur une approche descriptive. Les unités dégagées fournissent de bons outils pour la description structurelle des interactions et confèrent une grande puissance à l'analyse. Son potentiel reste encore aujourd'hui à exploiter dans la conception des systèmes interactifs, et tout particulièrement, dans la conception des systèmes de dialogue. On peut d'ores-et-déjà noter à ce sujet que, parmi les problèmes mentionnés par Schegloff (2006), les compétences des systèmes se concentrent surtout sur la sélection des mots et sur l'organisation de la structure globale. En revanche, l'interaction avec les systèmes automatiques actuels est d'une grande rigidité en ce qui concerne les prises et les cessions de parole, les séquences organisationnelles et les réparations de troubles. Ces points seront abordés au chapitre 2.

1.2.4 Conclusion

La théorie des actes de langage constitue les fondements de l'analyse pragmatique. Elle n'est pas contestée aujourd'hui bien qu'elle soit l'objet de nombreuses discussions et de certaines reformulations. La conclusion d'Austin demeure : la théorie de la signification ne peut pas ne pas tenir compte de ces faits.

Cependant, cette théorie n'est pas suffisante pour dresser un tableau complet du processus dialogique et des utilisations des énoncés. L'étude expérimentale de ces processus, à partir des travaux de Clark notamment (voir ci-dessous), a incité les psychologues à prendre en compte un point de vue plus large et à se focaliser sur « l'équipe » en communication. Les auteurs ont également accordé plus d'importance aux mécanismes représentationnels qui agissent entre les partenaires et au processus global d'interaction qui les unit.

1.3 Théorie de la collaboration

La théorie de la collaboration est issue de travaux réalisés en continuité de la théorie des actes de langage. Le principal représentant en est Herbert H. Clark, qui a fait paraître de nombreuses études depuis les années 1970 autour de la notion de *savoir mutuel* (Clark & Marshall, 1978, 1981; Haviland & Clark, 1974, 1977) et à qui l'on doit le *modèle de contribution* (Brennan & Clark, 1996; Clark, 2002; Clark & Brennan, 1991; Clark & Krych, 2004; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986). Ces travaux ont également des

conséquences applicatives largement étudiées (*e.g.* Allwood, Traum & Jokinen, 2000; Brennan & Clark, 1996; Brennan & Hulteen, 1995; Bunt, 1994; Cahn & Brennan, 1999; Whittaker, Brennan & Clark, 1991) qui seront abordées dans les chapitres suivants.

1.3.1 Construction conjointe de la référence

Alors que dans la théorie des actes de langage l'attention se porte sur les énoncés produits par les interlocuteurs, ce qui renvoie directement aux individus et à des problématiques psychologiques de compréhension et de production, dans le cas de la théorie de la collaboration, au contraire, l'attention se tourne vers le processus de communication et renvoie à un point de vue psychosocial (Krauss & Pardo, 2004). L'acte d'énonciation est étudié en tant qu'acte de référence, qui est vu comme un acte bilatéral impliquant à la fois le locuteur et l'auditeur (Clark & Krych, 2004). Une référence peut occuper plusieurs tours de parole. Elle implique au moins une proposition du locuteur et une sanction du destinataire. Ainsi, l'énoncé est inclus dans un processus plus général de communication, vu comme une *activité conjointe* (Clark, 1996). Le modèle correspondant est appelé *modèle sociocognitif du dialogue*. Dans ce modèle, le processus de conception des énoncés doit intégrer diverses contraintes.

Pour Clark et Wilkes-Gibbs (1986), la notion centrale est celle de *coordination* car elle renvoie au processus collaboratif nécessaire à la compréhension entre les interlocuteurs. L'expérience qu'ils proposent pour défendre ce point de vue est une *tâche de référence* issue des travaux de Krauss et Glucksberg (Krauss & Glucksberg, 1969, 1977). Elle est réalisée par deux participants, l'un jouant le rôle de *directeur* (*« director »*) et l'autre d'*exécutant* (*« matcher »*). L'objectif de l'exécutant est de ranger 12 figures (des *Tangram* chinois) dans l'ordre voulu à partir des indications du directeur. Dans l'expérience de Clark et Wilkes-Gibbs (1986), la tâche est réalisée six fois de suite avec la même série de figures. Les résultats montrent que le nombre de mots et de tours de parole nécessaires pour faire référence aux différentes figures diminue progressivement au cours de l'expérience, car les interlocuteurs construisent une base de connaissances communes au cours des échanges.

Le modèle descriptif du phénomène est un modèle par *acceptation* (*« acceptance »*) dans lequel le directeur propose un nom ou un groupe nominal qui peut être : (1) *accepté* par l'exécutant, pour indiquer au directeur qu'il attend sa proposition concernant cette référence, (2) *rejeté* par l'exécutant, qui peut préférer une autre désignation ou ne pas saisir la référence ou (3) *ajourné* par l'exécutant, qui peut réserver à plus tard son accord ou désaccord concernant la référence. Dans tous les cas, cette référence est réutilisable ultérieurement.

Dans ce modèle, l'effort des interlocuteurs n'est pas vu comme un effort individuel (comme dans Krauss & Glucksberg, 1977). Dans la mesure où le processus de construction de la référence est un processus collaboratif, la gestion de l'effort est partagée entre les partenaires. Les auteurs mettent donc en avant un *principe de moindre effort collaboratif* (« principle of least collaborative effort ») qui inclus les efforts partagés des partenaires. Ces

efforts dépendent (1) de la *pression temporelle*, (2) de la *complexité* de la référence et (3) de l'*ignorance* du locuteur quant à ce à quoi il réfère. La gestion de cet effort implique une gestion des réparations (de troubles) et des prises et cessions de parole, au sens de Schegloff, Jefferson et Sacks (1977).

Clark et Wilkes-Gibbs (1986), puis Clark et Brennan (1991) insistent sur la différence entre un supposé principe de *« moindre effort »* et celui de *« moindre effort collaboratif »*. En effet, dans le premier cas, il s'agirait d'une gestion personnelle de l'effort. Or, divers exemples permettent de montrer que l'effort individuel augmente parfois pour faciliter le processus collectif. Ce *'moindre effort'* engage donc bien l'équipe en communication. C'est pourquoi il est dit *'collaboratif'*.

1.3.2 Le modèle de contribution

A Contribution et terrain commun

A la suite de ces travaux, Clark et Schaefer (1989) proposent un modèle des contributions des interlocuteurs dans une conversation. Ce modèle propose de considérer toute *contribution* d'un participant à une conversation comme une partie d'un acte collectif de référence, lui-même construit sur l'organisation hiérarchique des *paires adjacentes* décrites par Sacks et al. (1974). La notion de *terrain commun* (*« common ground »*, Stalnaker, 1978) est alors utilisée pour désigner les présuppositions utilisées par les interlocuteurs pour construire leurs contributions. Pour Stalnaker :

> *« Le concept de présupposition implique que le locuteur part du principe que les membres de son audience présupposent tout ce que lui-même présuppose » (Stalnaker, 1978, p. 321, traduction libre).*

Ainsi, les *contributions* des participants sont *accumulées* par l'ensemble des partenaires et contribuent à leur *terrain commun*. Le processus de construction du terrain commun est nommé *« grounding ».* Son principe est celui du modèle d'acceptation :

(1) Une *phase de présentation* ;

(2) Une *phase d'acceptation*.

Un critère de grounding (*« grounding criterion »*) définit le niveau de compréhension nécessaire :

> *Le contributeur et les partenaires croient mutuellement que les partenaires ont compris ce que voulait dire le locuteur à un niveau suffisant pour le but actuel. (Clark & Schaefer, 1989, traduction libre)*

Sur cette base, différents types de contribution-réponse sont discriminés (Tableau 1-7). D'après Clark et Schaefer (1989), ce tableau permet de mettre en évidence une propriété cruciale : la *pertinence conditionnelle* (*« conditional relevance »*). En effet, si un locuteur produit la première partie d'une paire adjacente, il est pertinent et prévisible que son

interlocuteur produise la seconde partie de la même paire. Par là, l'interlocuteur B fournit trois éléments : (1) il croit avoir compris l'énoncé de A (dimension locutoire) ; (2) il a reconnu le type d'acte réalisé par A (dimension illocutoire) ; (3) il fournit une seconde partie en adéquation avec la première et affiche sa compréhension (dimension perlocutoire).

Tableau 1-7 : Les différents types de paires adjacentes

Type de paire adjacente		Exemple	
1ère partie	2de partie	Enoncé de A	Enoncé de B
Question	***Réponse***	Où est Connie ?	Au magasin.
Requête	***Conformité / Refus***	Passe-moi le tournevis s'il te plaît.	[B passe le tournevis]
Requête	***Acceptation / Rejet***	Passe-moi le tournevis s'il te plaît.	D'accord.
Proposition	***Acceptation / Rejet***	Voici votre change.	[B le prend]
Offre	***Acceptation / Rejet***	Voudriez-vous un café ?	Oui, merci.
Invitation	***Acceptation / Rejet***	Venez dîner ce soir.	D'accord.
Excuse	***Acceptation / Rejet***	Désolé.	Pas de problème
Remerciement	***Acceptation / Rejet***	Merci	De rien
Evaluation	***Accord / Désaccord***	Ce film était terrible.	Oui, tout à fait.
Compliment	***Accord / Désaccord***	Ton nouveau manteau est magnifique.	Oui, il est bien.
Convocation	***Réponse***	Hey, Ben.	Oui ?
Salut	***Salut***	Bonjour, Ben.	Bonjour, Anne.
Adieu	***Adieu***	Au revoir	Au revoir

Enfin, la réaction du destinataire dans la phase d'acceptation peut être plus ou moins marquée. Les auteurs identifient cinq niveaux : (1) l'*attention continue* envers le locuteur, (2) l'*initiation de l'énoncé pertinent suivant*, (3) la reconnaissance (« oui »…), (4) la *démonstration de compréhension*, (5) la *reprise* de tout ou partie de l'énoncé. Par principe, le destinataire est censé fournir une réaction suffisamment marquée.

B Indices et niveaux de coordination

Dans ses écrits plus récents (*e.g.* 1996; 2002; 2004), Clark mentionne qu'une variété d'actes est nécessaire à la coordination des partenaires en communication. Pour rendre compte du phénomène, Clark (2004) propose la coexistence de deux systèmes :

(1) Un *'système primaire'*, concernant le *contenu* proposé dans le signal, et ;

(2) Un *'système collatéral'*, assurant la *performance*[1] du discours, soit sa réalisation.

[1] Le terme « performance » est compris ici dans sa signification anglo-saxonne. To perform = accomplir, réaliser. Performance = représentation théâtrale, interprétation, réalisation.

En vertu de ces deux aspects, l'auteur précise que tout signal est composé d'une face de contenu et d'une face de performance, toutes deux nécessaire à la réussite de l'acte :

Signal = *contenu* + *affichage* (« content » + « display »)

La performance du locuteur peut être décrite sur la base d'un jeu d'indices permettant d'identifier : le *producteur* (locuteur), le *receveur* (destinataire), le *moment* (temps), la *location* (lieu) et le *contenu* du signal. A partir de ces indices, l'auteur montre qu'il est possible de qualifier les contributions des partenaires sous leurs différents aspects. Les partenaires travaillent ensemble et se coordonnent dans l'utilisation du langage du point de vue à la fois du *terrain commun* (« grounding ») et des *indices de performance*. Les signaux collatéraux pouvant être échangés sont :

- Les *insertions* : commentaires interposés ;
- Les *modifications* : marqueurs d'essai, intonations des voyelles, prolongations ;
- Les *juxtapositions* : remplacements de mots en cas d'erreur, répétitions, superpositions ;
- Les *concomitances* : mouvements de regard, expressions du visage, gestes du doigt.

Clark (2002) montre également que les *hésitations* et les *interruptions* (désignées, en anglais, par un terme unique « disfluencies », qui donne le néologisme français « disfluence », parfois utilisé) sont moins problématiques qu'on ne le pense généralement, et qu'elles peuvent être utilisées dans un but stratégique visant la coordination. Le problème ne serait pas tant la planification globale, préalable à l'énoncé, que sa planification en temps réel. L'auteur identifie quatre niveaux de coordination (*Cf.* Tableau 1-8) et indique que les stratégies d'utilisation des disfluences facilitent cette coordination. Il mentionne plusieurs stratégies tout en précisant que cette liste n'est pas exhaustive :

(1) *Signaler l'initiation de la parole*. Produire le début d'un énoncé puis marquer une pause permet de capter l'attention.

(2) *Respecter une prosodie idéale*. Il est possible de marquer un mot manquant ou d'interrompre et recommencer un énoncé mal construit pour produire une prosodie en accord avec les attentes du destinataire.

(3) *Signaler une intention de suspendre sa parole*. Il est, par exemple, possible de prolonger la dernière syllabe avant l'interruption.

(4) *Signaler une intention de différer*. Pour cet usage, le signal le plus fréquent en français est le 'euh'. Le délai d'attente peut être signalé par de légères différences, par exemple 'euhm'. Ces signaux s'adaptent évidemment aux circonstances particulières de l'énoncé.

(5) *Avant de suspendre sa parole, signaler quel type d'expression suivra*. Par exemple, dans « je ne serais pas surpris que… », l'auditeur s'attend à recevoir une prédiction du locuteur.

(6) *Signaler l'intention de réviser une expression*. C'est le cas lors de l'emploi d'une expression telle que : « je veux dire ».

Il y a donc une grande richesse dans l'emploi des disfluences qui assure la coordination entre les partenaires aux différents niveaux. Clark (2002) module le point de vue selon lequel ces signaux indiquent des difficultés de planification. Il fait remarquer que ces formes relèvent d'un emploi conventionnel et indique qu'elles devraient plutôt être considérées comme la solution aux problèmes de planification.

Tableau 1-8 : Les niveaux de coordination

Niveau 1	*L'auditeur doit suivre les vocalisations du locuteur pendant qu'il vocalise.*
Niveau 2	*L'auditeur doit tenter d'identifier les expressions présentées par le locuteur pendant qu'il les présente.*
Niveau 3	*L'auditeur doit tenter de comprendre ce que le locuteur veut dire pendant qu'il parle.*
Niveau 4	*L'auditeur doit examiner les projets communs proposés par le locuteur pendant qu'il les propose.*

On constate ici, sur la base de l'analyse conversationnelle, que Clark a pu détailler quels indices sont observables dans la réalisation d'un acte discursif. Il montre ainsi comment la signification est construite progressivement entre les partenaires et qu'elle n'est pas donnée à l'avance telle une définition du dictionnaire (Brennan & Clark, 1996). Cette construction repose sur une structure d'échanges fine, constituée d'énoncés adaptables.

1.3.3 S'adresser à l'auditoire

Le fait de s'adresser à un auditoire impose de prendre en compte une série de contraintes de façon à s'assurer de la bonne compréhension du message. La notion de *conception d'énoncé pour l'audience* (*« audience design »)* est destinée à rendre compte de ces contraintes et de leur intégration dans un comportement conversationnel coopératif (Clark & Carlson, 1982; Grice, 1975). L'analyse de la prise en compte de ces contraintes par le locuteur repose sur le même paradigme que les tâches de référence (tâches d'appariement, réalisées par un exécutant selon les indications d'un directeur), mais peut intégrer plusieurs exécutants auxquels des rôles différents sont attribués dans la communication. Ces situations permettent de rendre compte de la production d'énoncé sous l'angle des divers effets produits, c'est-à-dire, sous l'angle des fonctions de l'énoncé.

A L'acte de langage et les auditeurs

Clark et Carlson (1982), fournissent une analyse détaillée du phénomène de prise en compte de l'audience. Ils proposent en premier lieu, de distinguer quatre rôles fondamentaux dans la communication :

(1) Le *locuteur* ;

(2) Le d*estinataire ;*

(3) Le *participant ;*

(4) L'a*uditeur*.

Lorsqu'un *locuteur* produit un énoncé vers un *destinataire*, le *participant* prend part à l'acte (le *locuteur* en attend une certaine réponse) alors que l'*auditeur* n'y prend pas part (il occupe un rôle passif).

D'après Clark et Carlson (1982), la production d'un acte de parole impose d'assigner des rôles aux personnes présentes, soit par une désignation verbale explicite, soit par un geste. Ils précisent que les gestes ont l'avantage d'être des actes publics facilement reconnaissables par tous les participants. L'exemple utilisé est emprunté à Shakespeare :

Othello, à Desdemona, face à Iago et Roderigo : « Viens, Desdemona ».

Sur la base de cet exemple, les auteurs avancent trois hypothèses (l'analyse proposée repose sur la *« m-intention »* de Grice, 1969, Cf. Tableau 1-3). Tout d'abord, Clark et Carlson (1982) proposent de considérer que l'acte illocutoire dirigé vers la *destinataire* (Desdemona) est différent de l'acte illocutoire dirigé vers les *participants* (Iago et Roderigo). Ensuite, ils proposent de considérer que l'acte illocutoire fondamental est l'acte par lequel le locuteur informe tous les participants qu'il réalise un acte illocutoire dirigé vers la destinataire ; et ils nomment *'informatif'* (*« informative »*) cet acte fondamental. Enfin, ils proposent de considérer que tout acte illocutoire dirigé vers un destinataire (nommé '*assertif' : « assertive act »*) est réalisé par le moyen d'un *informatif*. Ces hypothèses, non formulées dans la théorie générale (théorie des actes de langage), permettent alors d'expliquer comment le locuteur communique ses intentions dans la conversation ordinaire, comment il peut produire des actes indirects, désigner son destinataire, ou encore, faire participer le public ou des témoins.

B Formalisation des actes informatif et assertif

L'objectif des auteurs est de proposer une écriture formelle des sous-intentions du locuteur (Figure 1-2). L'informatif est représenté sur la première ligne par trois arguments : le locuteur (Othello, noté « O »), les participants (Desdemona, Iago et Roderigo, notés « D & I & R ») et l'acte illocutoire de second rang (noté « Requête »), c'est-à-dire l'assertion en elle-même. Cette requête, adressée à Desdemona est représentée sur la seconde ligne. Elle est également composée de trois arguments : le locuteur (Othello, noté « O »), la destinataire (Desdemona, notée « D ») et l'acte illocutoire assertif (noté « D vient avec O »).

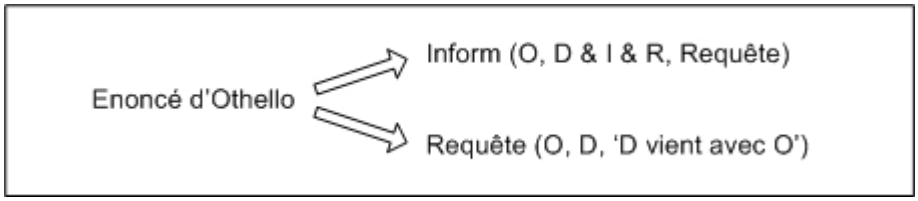


Figure 1-2 : Décomposition de l'acte illocutoire

A l'aide de cette décomposition, les auteurs dégagent la forme logique de divers exemples et montrent comment des significations complexes peuvent être mises en évidence. Par

exemple, dans une soirée, la femme d'un convive peut rappeler à son mari qu'il a un rendez-vous le lendemain matin pour, à la fois, lui indiquer qu'il est temps de partir, indiquer à un autre convive qu'il n'est pas nécessaire de lui servir à boire, et à un troisième, que la blague qu'il vient de raconter n'était pas drôle. Ces analyses ne seront pas détaillées ici.

Clark et Carlson (1982) concluent qu'il s'agit là d'un ajout fondamental à la théorie des actes de langage dans la mesure où cette formalisation permet d'analyser en profondeur la dimension illocutoire des actes. Cela va en effet dans le sens de la théorie des valeurs illocutoires voulue par Austin. Mais ils notent que l'analyse d'un acte illocutoire standard, n'impliquant qu'un locuteur et un auditeur, n'est pas modifiée. Ils n'exploitent pas cette distinction entre *informatif* et *assertif* dans le contexte spécifique de la production d'énoncé.

C Terrain commun en production et en compréhension

Diverses études ont été réalisées dans le but de mettre en évidence la manière dont les interlocuteurs utilisent les connaissances du terrain commun (Clark & Krych, 2004; Horton & Gerrig, 2005a, 2005b; Horton & Keysar, 1996; Krauss & Fussell, 1989; Schober & Clark, 1989). Schober et Clark (1989) montrent d'abord que le destinataire des explications d'un directeur, en interaction avec lui, est plus apte à ranger des figures qu'un auditeur placé en arrière plan et ne participant pas aux interactions. Le rôle du processus de *grounding*, *i.e.* des interactions entre interlocuteurs, est donc évident. A partir de ce résultat, d'autres études sont vouées à vérifier si les connaissances issues du *terrain commun* sont utilisées de façon systématique lors des processus de production et de compréhension. Concernant la compréhension, Keysar et Paek (1993) montrent que le destinataire tend à interpréter la référence d'un locuteur en fonction des référents dont il dispose, sans prendre en compte leur accessibilité pour le locuteur. L'interprétation serait donc réalisée de façon *égocentrique* et elle serait ajournée en cas d'échec. De retour à la problématique de la production des énoncés, Horton et Keysar (1996) opposent un modèle de *conception initiale*, dans lequel les connaissances du terrain commun seraient prises en compte dès le premier énoncé, à un modèle de *surveillance et ajustement* (*« monitoring and adjustment »*), dans lequel une expression est proposée sur un mode égocentrique (à nouveau) et ajustée au cours des interactions. Les résultats montrent que les connaissances initiales peuvent être prises en compte dès le premier énoncé quand la tâche est réalisée sans pression temporelle. En revanche, quand les circonstances se font pressantes, le locuteur se repli sur le mode égocentrique.

Les auteurs concluent que la prise en compte des connaissances communes lors de la conception pour l'audience n'est pas inhérente aux routines de la conception d'énoncé. Certains énoncés sont conçus correctement sans référence au terrain commun. En cas de violation du terrain commun, ils sont révisés interactivement. Ce fonctionnement est cohérent avec le processus initialement décrit par Clark et ses collaborateurs (Clark & Marshall, 1981; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986). Ce résultat met en évidence le rôle important joué par des processus de bas niveau dans la conception routinière des énoncés.

1.3.4 Conclusion

Les apports de Clark et de ses collaborateurs (ainsi que de ceux qu'ils ont influencés) résident dans la description fine du processus collaboratif et dans les démonstrations expérimentales qu'ils proposent pour défendre chaque notion.

Cependant, les processus mis en avant se rapportent à l'*équipe* engagée dans le processus collaboratif. De ce fait, ces auteurs accordent moins d'importance aux processus de production et de compréhension des énoncés qui jouent à l'échelle de l'individu. Ainsi, la mise en évidence des aspects performatifs du discours n'est pas suffisante pour tous les auteurs.

1.4 Hiérarchie des processus liés à la conception des énoncés

Le modèle de contribution tel qu'il est formulé par Clark et ses collaborateurs suppose des traitements cognitifs coûteux qui seraient dédiés à la construction d'un modèle des connaissances du (ou des) partenaire(s). Certains auteurs (*e.g.* Pickering & Garrod, 2004) souhaitent donc mieux décrire les processus individuels qui sous-tendent ce fonctionnement pour préciser les rôles joués par les différents types de processus.

1.4.1 Le rôle des automatismes

La conversation est le site essentiel du discours (Garrod & Pickering, 2004). Pour ces auteurs, si cette activité est facile pour tout-un-chacun, c'est que les mécanismes de traitement interactif reposent sur un processus d'alignement automatique des représentations linguistiques entre les partenaires. Ce point de vue est développé sous le nom de *théorie mécaniste du dialogue* (Pickering & Garrod, 2004). Les auteurs proposent tout d'abord d'établir une différence entre la notion de *coordination*, qui concerne les actions des participants, et la notion d'*alignement*, qui porte sur les représentations des participants. Six points sont ensuite argumentés. Ils sont présentés dans le Tableau 1-9.

Tableau 1-9 : Les six arguments du modèle d'alignement interactif

(1) L'alignement des modèles de situation est la base d'un dialogue réussi.
(2) L'alignement des modèles de situation est effectué grâce à un mécanisme d'amorçage automatique.
(3) Le même mécanisme produit l'alignement de la représentation aux autres niveaux (lexical et syntaxique).
(4) Les interconnexions entre niveaux impliquent que l'alignement à un niveau conduit à l'alignement aux autres niveaux.
(5) Un autre mécanisme primitif permet aux interlocuteurs de réparer interactivement les représentations qui ne sont pas correctement alignées.
(6) Des stratégies plus sophistiquées et coûteuses de modélisation des états mentaux de l'interlocuteur ne sont requises qu'en cas d'échec des mécanismes primitifs à produire l'alignement.

Ces propositions permettent de souligner la souplesse du système cognitif. Les processus de production et de compréhension du discours sont habituellement vus comme une série de traitements modulaires (Fodor, 1983; Fodor, Bever & Garrett, 1974) et successifs (Levelt, 1989; Levelt, Roelofs & Meyer, 1999). Notamment, pour Levelt, les différents traitements sont présentés comme allant « de l'intention à l'articulation » (sous-titre de l'ouvrage de 1989). Cette vision repose principalement sur les temps de traitement correspondant aux différents niveaux de représentation. Ainsi, chaque module de traitement génèrerait, en sortie, un certain niveau de représentation, qui peut alors être utilisé en entrée du module suivant pour construire un autre niveau de représentation, etc. Ces traitements en chaîne se poursuivraient jusqu'à obtenir l'articulation d'un son, qui peut ensuite faire l'objet des traitements inverses, pour décodage, chez le destinataire.

Pickering et Garrod (2004) ne nient évidemment pas l'existence des différents niveaux de représentation, mais ils rejettent l'idée d'une structure aussi rigide. Pour eux, certains effets d'activation de représentation sont liés au contexte de la communication plutôt qu'à la chaîne modulaire en elle-même. Les différents niveaux de représentation sont interprétés indépendamment les uns des autres par le destinataire. A travers chacun de ces niveaux les liens entre les interlocuteurs sont nombreux (Figure 1-3). Cette vision est contraire à celle d'une transmission autonome, qui n'admet de lien que via la chaîne sonore, au niveau phonétique (lien le plus bas de la figure). De plus, l'activation des représentations reposant sur un mécanisme d'amorçage permet d'expliquer comment les interlocuteurs s'influencent mutuellement au cours des échanges et comment ils maintiennent des représentations équivalentes grâce aux interactions. Cela explique en particulier pourquoi il est si fréquent que le destinataire d'un message termine lui-même l'énoncé que lui adresse le locuteur.

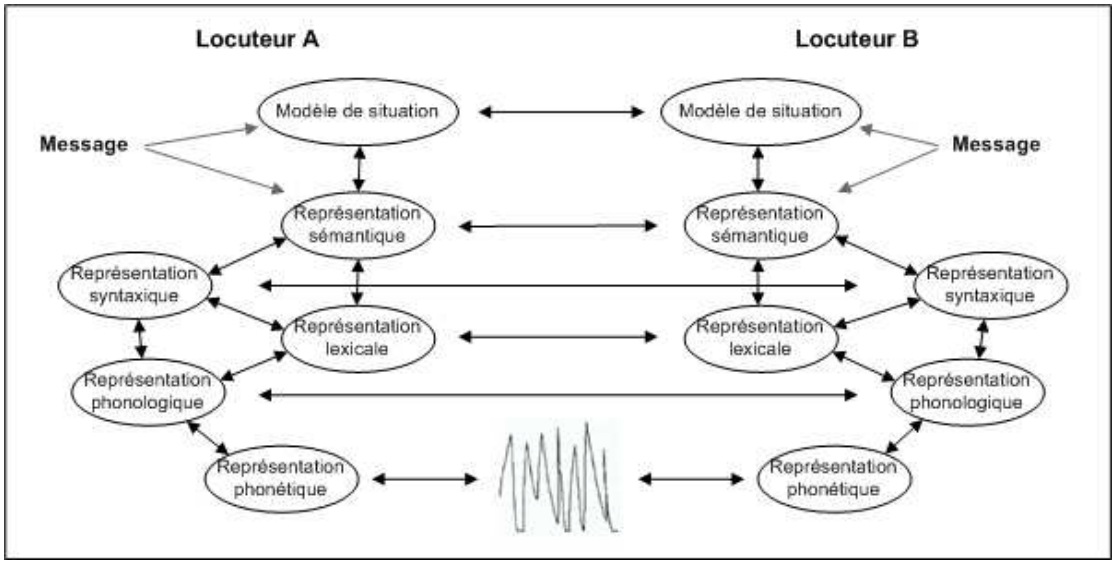


Figure 1-3 : Modèle d'alignement interactif

Ainsi, le modèle d'alignement interactif incite à prendre en compte la souplesse des traitements grammaticaux. Il montre en particulier que les différents niveaux de représentation ont une indépendance relative. En faisant cette proposition, Pickering et Garrod (2004) nous offrent une possibilité de présupposer que des effets peuvent être

produits à partir de chaque niveau, indépendamment du (ou parallèlement au) contrôle des autres niveaux, et du niveau supérieur en particulier. Ainsi, ils offrent une alternative à une vision strictement intentionnelle de la gestion de l'interaction. Il ne s'agit pas nécessairement d'un argument nouveau dans la mesure où celui-ci a été avancé par plusieurs auteurs (Brooks, 1991; Glenberg, 1997; Lakoff, 1987; Putnam, 1983; Varela, Thompson & Rosch, 1993). Il s'agit en revanche d'une intégration nouvelle de l'argument dans un modèle psycholinguistique. Par là, cette approche renouvelle l'intérêt pour l'étude des processus de bas niveau.

1.4.2 Le rôle des processus de mémoire

Récemment, Horton et ses collaborateurs (Horton, in press; Horton & Gerrig, 2005a, 2005b; Horton & Spieler, 2007) se sont focalisés sur la manière dont les locuteurs s'adaptent aux contraintes qui régissent la conception de leurs énoncés. L'objectif premier de leurs études est de chercher à pondérer et/ou à animer les différents processus psychologiques qui interviennent dans la construction de l'énoncé.

Horton et Gerrig (2005a) présentent les phénomènes intervenant dans la conception des énoncés comme les caractéristiques émergentes des processus ordinaires de mémoire. Ces auteurs souhaitent fournir des résultats permettant de décrire la conception d'énoncés pour l'audience d'une façon cohérente à la fois avec les modèles de production du langage et les *processus cognitifs ordinaires*, plus généraux, émanant de la psychologie cognitive. Pour cela, ils ajoutent des contraintes catégorielles au protocole de Schober et Clark (1989). Ils font arranger les cartes, sous les indications du même *directeur*, par deux *exécutants* différents, dans des séries réalisées successivement. Les cartes à arranger représentent des animaux. Chaque épreuve consiste à ranger quatre cartes. Les séries sont constituées de quatre épreuves successives, soit 16 cartes au total. Avec la moitié des triades, chaque série de quatre cartes est composée d'une seule catégorie (quatre chats, quatre poissons, etc.). Cette condition est nommée *orthogonale*. Avec l'autre moitié, les catégories sont croisées entre les deux exécutants de façon à compliquer l'association catégorie-exécutant ; condition *croisée*. Les résultats montrent des temps de planification des énoncés plus longs pour les cartes nouvelles et un effort supplémentaire nécessaire dans la condition croisée. Par ailleurs, l'adaptation du contenu des énoncés au destinataire est plus évidente dans la condition orthogonale, indiquant des difficultés d'adaptation dans la condition croisée. En vertu de ces résultats, les auteurs valident l'importance des processus de mémoire dans le processus de conception des énoncés pour l'audience.

Mais d'autres travaux (Bard et al., 2007) montrent qu'un modèle basé sur la *charge cognitive*, vue comme la charge pesant sur les processus mémoriels impliqués dans l'adaptation aux connaissances du locuteur, prédit moins bien la performance qu'un modèle de *responsabilité partagée*, incluant le processus de grounding. Ainsi, même s'ils jouent un rôle important, les processus ordinaires de mémoire ne sont pas suffisants pour rendre compte du phénomène

de conception des énoncés dans son ensemble. Horton et Gerrig (2005b) vont également dans ce sens. Ils distinguent un processus de *formation du message* (*« message formation »*), portant sur la forme du message adressé au destinataire, d'un autre processus dédié à l'*évaluation du partage communautaire* (*« commonality assessment »*) du contenu, lié à la problématique de la conception pour l'auditoire (*« audience design »*). Ils insistent sur l'évaluation du partage communautaire et renvoient à ce sujet à l'article de Clark et Marshall (1981). Pour ces derniers, le problème des connaissances mutuelles relève d'une triple coprésence. La *coprésence physique* réfère à l'environnement matériel immédiat. La *coprésence linguistique* réfère à la conversation entre les interlocuteurs. Et l'*adhésion communautaire* réfère à l'arrière plan socioculturel. Mais malgré ces efforts de catégorisation des différents processus, les résultats de Bard et al (2007) semblent indiquer que le processus de *formation du message* ne peut être considéré tout-à-fait séparément de l'utilisation de ces connaissances liées à la triple coprésence.

1.4.3 Conclusion

Globalement, la question qui se pose dans ces études, et dans la notion d'*audience design*, est celle de la prise en compte d'autrui et de ses connaissances. Ces travaux montrent que l'empathie n'est pas la règle, ou du moins qu'elle n'est pas systématique, et que le mode de fonctionnement primordial est le mode égocentrique. L'empathie n'apparaît que lorsque les interlocuteurs ont la possibilité d'exploiter les connaissances communes, si elles existent et si les contraintes liées à l'activité le permettent. Mais ce processus n'est pas primaire. De façon plus primordiale, dans l'interaction, la forme des énoncés conçus repose sur les interactions qu'ils engendrent. Celles-ci sont décrites d'un point de vue psychosocial par le *'modèle d'alignement interactif'* qui occupe un rôle central. L'approche de Pickering et Garrod (2004) montre en effet que les automatismes jouent un rôle primordial ; et les résultats de Bard et al. (2007) indiquent qu'une focalisation sur le processus mémoriel est insuffisante et que le processus interactif doit être pris en compte en premier lieu.

Ainsi, bien qu'ils permettent le contrôle de l'activité, les processus de haut niveau, par exemple, la mémoire, ne sont pas des processus qui *englobent* les processus de plus bas niveau, mais qui, au contraire, y *sont englobés*. Cette idée a été traduite par Brooks (1991) sous le nom de *« subsomption »* et son caractère essentiel est désormais admis en intelligence artificielle (*e.g.* Wooldridge, 2002).

Mais, si les études prouvent le rôle primordial des automatismes, la description fine de leur(s) utilisation(s) en contexte reste un défi méthodologique et technique. Ce type de description permettrait de montrer quels automatismes organisent quelles interactions et, probablement, de montrer leur prise en compte et leurs effets sur le niveau des connaissances (quelles inférences sont produites à partir des automatismes ?).

1.5 La question de la pertinence

La question de la pertinence était déjà évoquée par Austin (1962), en termes d'action. Elle est évoquée dans les mêmes termes, plus récemment, par Schegloff (2006). Cependant, aucun de ces deux auteurs ne cherche à préciser plus formellement la notion.

Les maximes de Grice (1975) et la pertinence conditionnelle de Clark et Schaefer (1989) apportent un point de vue fonctionnel qui se rapproche, comme le remarquent les auteurs, de la dimension illocutoire austinienne. Elles agissent en effet comme des connaissances conventionnelles mises en pratiques par les interlocuteurs dans l'interaction. Telle qu'elle est définie par Clark et Schaefer, la pertinence d'un énoncé dépend de sa proximité avec la suite perlocutoire attendue par convention (à la suite de l'énoncé précédent, ou de l'action en cours). Les maximes de Grice expriment quant à elles les règles de construction internes de l'énoncé. Globalement, ces descriptions conduisent à une définition de la notion de pertinence en relation au référentiel culturel.

1.5.1 Théorie de la pertinence

Une proposition différente a été avancée par Sperber et Wilson (1989). Leur théorie est intéressante à plusieurs égards et elle a, de ce fait acquis une renommée qui impose de la présenter (même rapidement). Mais cette présentation va seulement permettre d'expliquer pourquoi elle doit être écartée.

Sperber et Wilson (1989) proposent une critique des principes gricéens. Ils souhaitent dépasser l'approche descriptive des maximes pour proposer une théorie explicative. Pour eux, l'énoncé est un acte d'ostension (un acte volontaire) qui véhicule deux niveaux d'information. (1) L'information *directe* est l'*intention communicative*. Elle indique au destinataire que le communicateur souhaite attirer son attention sur un phénomène. (2) L'information *indirecte* est l'*intention informative*. Elle précise le *vouloir-dire* du locuteur concernant le phénomène. Elle est indirecte car elle se greffe sur l'intention communicative et parce qu'elle suppose une interprétation de la part du destinataire. Cette distinction est très proche de celle de Clark et Carlson (1982) entre acte informatif et assertif.

Le *principe de pertinence* est substitué au *principe de coopération*. Il stipule que, lorsqu'il prend la parole, le locuteur s'engage à proposer un *'acte ostensif'* pertinent. La *pertinence* de cet acte est analysée sur la base des effets contextuels qu'il entraîne. Elle est définie par le rapport entre l'*effet* contextuel produit et l'*effort* nécessaire pour le traiter. D'après le principe de pertinence, la signification qui sera retenue de l'énoncé est celle qui produit l'*effet* contextuel le plus grand, pour l'*effort* le plus faible. Allwood (1995) synthétise cette proposition en expliquant qu'elle permet de réduire les maximes gricéennes à une seule d'entre elles, la maxime de relation.

Aucun auteur ne conteste l'objectif affiché par Sperber et Wilson d'évoluer vers une théorie plus explicative. Mais, bien que la *théorie de la pertinence* ait eu une influence, par exemple, sur la conception d'outils de communication permettant de reproduire certains éléments contextuels (*e.g.* Dumazeau, 2005; Zouinar, 2000), elle ne fait pas l'objet d'un consensus. Par exemple, Caron (1989) note que Sperber et Wilson font une scission trop radicale entre les aspects pragmatiques et linguistiques du discours. Mais surtout, un modèle des actions d'un individu en contexte naturel ne peut se résumer à une équation entre deux variables. Une analyse complète des actions est nécessaire (Schegloff, 2006).

1.5.2 Le point de vue d'Allwood

Allwood (1984; 1995) propose une analyse plus complète des actions. Pour lui, la pertinence est un concept relationnel. Elle dépend d'une *relation moyen-fin* qui opère dans la situation.

Pour établir son point de vue, l'auteur rappelle d'abord plusieurs éléments. Pour lui, la communication doit être vue comme une activité multi-niveaux (physique, biologique, psychologique et social) dans laquelle les niveaux s'entremêlent et fournissent des opportunités et des contraintes. L'auteur rappelle l'analyse qu'il a faite de la notion de coopération (Allwood, 1976) à la suite des propositions de Grice (1975) et qui propose de différencier quatre plans : (1) la *considération cognitive,* par les partenaires, du problème évoqué, (2) la *proximité des buts* des partenaires, (3) les *considérations éthiques* entre les partenaires et (4) la *confiance* réciproque. Ces plans d'analyse sont proposés comme des points de vue qui permettent d'éclairer la situation étudiée. Mais l'objet d'analyse que pose l'auteur est l'*activité* déployée par les interlocuteurs. Son unité de référence est l'énoncé (*« contribution »* et *« utterance »*, les deux termes sont revendiqués) et l'organisation des séquences en fournit la structure. Chaque énoncé est décrit selon sa forme de surface, nommée *caractéristiques d'expression* (*« expression features »* : geste ou oral, structure acoustique et syntaxique), et son contenu (*« content features »*) qui véhicule un sens plus ou moins explicite. L'auteur distingue trois fonctions associées au contenu : (1) une *fonction de gestion propre* de ses énoncés par le locuteur, (2) une *fonction interactive* qui permet à l'auditoire de réagir et (3) les *autres fonctions communicatives* telles que l'assertion, le questionnement, la promesse, etc. Il précise qu'un énoncé peut être mono ou multifonctionnel.

L'analyse ne s'arrête pas là. L'acte communicatif est encore vu selon deux dimensions qui permettent de distinguer ses effets : (1) la *dimension expressive* permet au locuteur d'exprimer une attitude auprès de l'interlocuteur et (2) la *dimension évocatrice* permet d'évoquer une réaction chez l'interlocuteur (une croyance). Cette dernière distinction permet à l'auteur d'en arriver à la notion d'obligation. En effet, le locuteur se doit d'être sincère et cohérent quant à ce qu'il exprime, et de prendre en considération les croyances de son interlocuteur. Mais également, le destinataire est obligé de produire certaines actions à sa suite, pour évaluer s'il veut poursuivre, s'il perçoit, comprend, etc. Les *obligations de considération* sont distinguées des *obligations de réponse*. Pour Allwood, ces aspects de

coordination sont les plus importants dans la communication car ils permettent d'assurer la cohésion sociale. Le caractère éthique de la relation prend ici une importance toute particulière. Finalement, tout énoncé est vu à travers son caractère explicite (implicite *vs* explicite), sa polarité (positif *vs* négatif), sa fonction de feedback (contact, perception, compréhension, acceptation) et son rôle de gestion interactive (séquençage et gestion des tours).

Ce n'est qu'à la suite de cette mise en ordre des fonctions de l'énoncé qu'Allwood en vient à définir la notion de pertinence. Il rappelle la nature relationnelle de cette notion, admet qu'elle est multiple, admet qu'elle a des degrés divers et qu'elle dérive du système de communication super-ordonné qu'elle caractérise. Ces précautions étant prises, quatre niveaux sont alors proposés pour décrire la pertinence d'un énoncé (Tableau 1-10).

Tableau 1-10 : Les niveaux de pertinence

Pertinence primaire	*Elle correspond à une évaluation positive ou négative des intentions évocatives de la contribution précédente, relativement aux intentions visées. Elle est communiquée explicitement ou implicitement.*
Pertinence secondaire	*Elle correspond au contact, à la perception et à la compréhension des contributions. Elle dépend de la continuité entre les contributions des partenaires et répond à la fonction évocative de la contribution précédente.*
Pertinence tertiaire	*Elle correspond directement à l'objet de l'activité. Elle dépend de la proximité de la contribution avec le but et le contexte sémantique.*
Pertinence Quaternaire	*Elle correspond à des contributions portant plus spécifiquement sur des aspects contextuels. Elle dépend de la prise en compte des éléments extérieurs*

Pour Allwood (1995), cette description sur plusieurs niveaux permet de capturer les aspects importants de la pertinence dans le flux de la communication, et elle permet dans le même temps de capturer la notion de cohésion. Pour l'auteur, cette analyse montre que la considération mutuelle dans la communication repose sur des maximes d'action rationnelle et motivée telles qu'il en existe dans *« l'obligation de réponse »*.

1.6 Conclusion

1.6.1 Bilan : qu'est-ce qu'un énoncé ?

Cette synthèse d'Allwood (1995) nous offre un point de vue très complet sur l'énoncé, sa structure, ses fonctions et les moyens de les qualifier. C'est la structure interactionnelle qui prime dans ce système car elle est le liant par les moyens duquel les individus entrent en relation. C'est sur cette base que les processus cognitifs supérieurs peuvent se mettre en place (Bard et al., 2007).

Ainsi, l'énoncé est un acte ou un ensemble d'actes produit(s) par un individu avec le but de provoquer différents effets sur un auditoire. Cet énoncé est interprété au travers de conventions par cet auditoire dans lequel plusieurs rôles peuvent être distingués. En fonction du rôle attribué à chacun les interprétations qu'il fera seront différentes et produiront chez lui des effets différents. Ces effets portent (1) sur les connaissances que le locuteur a souhaité transmettre, (2) sur la relation et les positionnements sociaux qu'il induit entre les participants et (3) sur la (ou les) réponse(s) obtenue(s). Ces éléments seront synthétisés et précisés au chapitre 5 pour en dégager les questionnements de la thèse.

Ainsi, l'énoncé ne transmet pas des « informations » comme on peut l'entendre au sens classique de la théorie de la communication (Shannon, 1948). Les effets produits sont plus larges, ce qui a été exprimé par Michel Foucault (1976, p. 133) – quoique dans une perspective différente – par la phrase suivante : *« Le discours véhicule et produit du pouvoir. »* C'est ce que montrent les différents effets identifiés : pouvoir d'amener l'autre à construire des connaissances, pouvoir de créer du positionnement social entre les individus, pouvoir de provoquer chez l'autre des réactions. Tout cela est très différent en réalité d'une simple « *transmission d'information »*.

Enfin, au travers de ces effets et en vertu de l'expérience commune des participants, ils *construiront* un ensemble de connaissances communes qu'ils pourront réutiliser dans la suite de l'interaction et dans des interactions ultérieures.

1.6.2 Perspectives

L'approche pragmatique tend aujourd'hui à être utilisée comme le paradigme de référence pour l'analyse et le développement des systèmes de *dialogue homme machine* (Allwood et al., 2000), problème qui est abordé aux chapitres 2 et 3.

D'après un visionnaire du domaine (Goffman, 1987), les pistes d'amélioration de ces théories passent par l'*analyse micro-fonctionnelle*. Cette analyse recouvre les préoccupations de plusieurs auteurs qui cherchent à préciser le fonctionnement des interactions (*e.g.* Clark, 2002), la prise en compte d'autrui dans l'interaction (*e.g.* Brennan & Clark, 1996; Clark & Krych, 2004) ou le rôle des processus cognitifs ordinaires permettant la conception des énoncés (Bard et al., 2007; Holtgraves, 2008; Horton & Gerrig, 2005a). La méthodologie d'étude de ces problématiques n'est pas uniforme, ce qui ne facilite pas la comparaison entre les études. Dans ce cadre, l'étude systématique des effets perlocutoires et des suites perlocutoires provoqués par les énoncés en fonction des conventions illocutoires et des marqueurs locutoires utilisés reste un objectif à atteindre pour l'approche expérimentale. Le dialogue homme machine offre un paradigme intéressant dans ce but car il permet de préciser certains questionnements et parce qu'il offre un paradigme d'étude spécifique.

Cette littérature souligne la souplesse fonctionnelle des interactions et la finesse que cela suppose lors de l'analyse. Dans la perspective de la conception de systèmes automatiques, cette souplesse s'impose aux concepteurs. Sans elle, un système de dialogue n'est rien de

plus, en fait, qu'une machine à délivrer des messages face à laquelle le dialogue reste fictif. La nuance est grande et l'utilisateur s'en rend compte rapidement. En particulier, il est nécessaire pour dialoguer de savoir reconnaître les fonctions des interventions. Par exemple, il arrive régulièrement qu'un énoncé soit inféodé à un autre (Kerbrat-Orecchioni, 1995), *i.e.* qu'il soit hiérarchiquement dépendant d'un autre. Cet acte ne constitue pas, alors, une intervention à proprement parler et ne doit pas être traité comme tel. Dans un système de dialogue automatique, la priorité du traitement réalisé par la reconnaissance vocale sur cet acte de langage pourrait, par exemple, être adaptée. Ainsi, ces principes d'analyse permettent de questionner le fonctionnement des systèmes automatiques.

Chapitre 2 Conception des systèmes de DHM

L'analyse pragmatique, présentée au chapitre 1, fournit le modèle de référence principal qu'utilise *l'ingénierie* pour la conception de *systèmes de Dialogue Homme Machine* (DHM). La conception s'appuie sur des techniques développées pour *formaliser l'activité* de dialogue et *implémenter des systèmes* doués de capacités dialogiques. Dans cette perspective, un corps de méthodes et d'études empiriques a été constitué de façon à stabiliser le processus de conception et les connaissances nécessaires au développement. C'est au cours de ce processus que sont conçus les énoncés du système.

Ce chapitre est d'abord centré sur l'aspect industriel car cet aspect est nécessaire pour définir les systèmes de dialogue. Suite à cette présentation, les développements théoriques permis, en psychologie, grâce à ces systèmes, sont présentés. Cette présentation permet de souligner les liens entre théorie et application.

2.1 Principes méthodologiques de l'Ingénierie Cognitive

L'ingénierie cognitive, ou *ingénierie des systèmes cognitifs*, renvoie simultanément au fonctionnement de la cognition humaine (pensée, résolution de problème, prise de décision) et aux concepts, méthodes et outils utilisés par l'ingénierie pour le développement de systèmes dédiés à l'assistance des humains (Rasmussen, Pejtersen & Goodstein, 1994). Dans ce domaine, l'étude de la cognition est donc soumise à un impératif de développement des systèmes techniques.

Le point de vue est général. Il porte sur la conception du système. Dans cette perspective, la discipline adopte un *modèle intégré de l'humain et du système* plutôt que de se centrer sur l'un ou l'autre des deux membres. Il n'y a pas non plus de point de vue spécifique sur la conception des énoncés. Ceux-ci sont intégrés au processus de fonctionnement global du système. Ce fonctionnement est envisagé sous l'angle de la conception écologique des systèmes d'information (Gibson, 1979) qui suppose la capacité humaine à lire, en environnement naturel, les propriétés fonctionnelles et les possibilités d'action. Dans cette vision, une notion fondamentale est celle de *couplage fonctionnel* (*Cf.* Figure 2-1, issue de Rasmussen et al., 1994, page 124). Cette notion renvoie à la fois à l'individu qui agit et au système dans lequel il agit. Elle peut être rapprochée du concept d'affordance (Gibson, 1979). L'efficacité de ce couple dépend de la compatibilité entre ses deux membres. La conception du système repose sur la performance cette équipe dans le contexte sociotechnique de l'activité. Cette dualité impose de faire converger les paradigmes des différentes disciplines académiques (psychologie, sociologie et sciences de l'ingénieur, notamment). La recherche

est dite *conduite par les problèmes* (*« problem-driven »*) et doit répondre aux impératifs de développement des systèmes. Dans ce cadre, le critère d'évaluation des modèles conceptuels consiste à savoir s'ils sont utiles et s'ils ont un pouvoir prédictif en contexte écologique (Rasmussen et al., 1994).

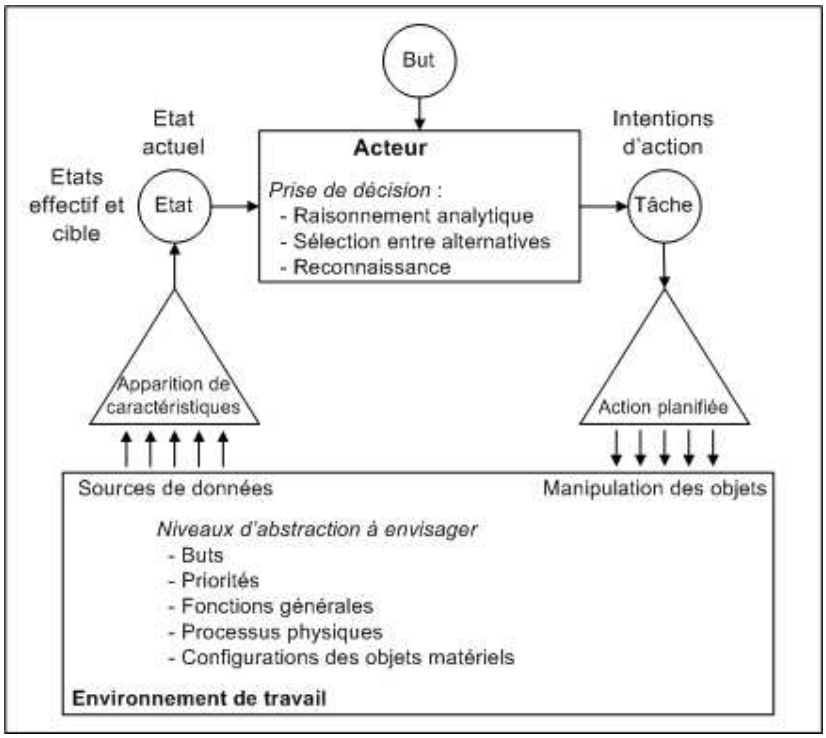


Figure 2-1 : Schématisation du couplage fonctionnel

Pour ces raisons, l'ingénierie cognitive s'oriente sur la méthodologie de conception des systèmes sociotechniques et s'organise autour des phases de planification, de conception et d'évaluation de ces systèmes. Cela suppose d'adopter une perspective fonctionnelle qui intègre et dépasse la perspective structuraliste, habituelle en sciences-humaines. Rasmussen et al. (1994) notent que la limite de l'analyse structurelle provient de l'absence de stabilité dans les réponses de l'acteur humain. Pour une structure donnée du système cognitif humain, il n'existe pas d'association stable entre les stimuli en entrée et les réponses en sortie. Il est donc impossible de proposer une structure hiérarchique globale agrégeant des relations de cause à effet étudiées isolément. Des modèles conçus isolément n'expliquent pas la variabilité totale du comportement car les humains s'adaptent aux caractéristiques fonctionnelles du système. Ils modifient leur comportement en fonction de leurs besoins.

Ainsi, l'analyse suppose de s'intéresser aux buts des utilisateurs et aux contraintes qui pèsent sur leur comportement face au système. Cette méthode d'analyse conduit à prendre en compte des *patterns comportementaux* qui apparaissent dans la continuité de l'activité. Ainsi, les variables ne peuvent plus être observées isolément. La synchronisation des *patterns situationnels* et des *patterns comportementaux* doit être mise en évidence. Elle renvoie au contrôle de la situation par l'acteur humain. Rasmussen et al. (1994, page 14) précisent que la notion de contrôle ne renvoie pas uniquement à la prise de décision consciente, qui suppose des opérations cognitives abstraites, mais également aux niveaux de contrôle

inférieurs, basés sur les règles et sur les automatismes, opérant au cours d'une activité continue. La notion de gestion (« *management* ») de l'activité permet alors de désigner cette fonction de contrôle opérant sur différents niveaux. Le succès de la modélisation et de la conception d'un système adaptatif dépend de la complétude des éléments pris en compte lors des phases d'analyse préalables au développement du système.

2.2 Initiation aux systèmes de dialogue homme machine

La conception des systèmes de dialogue est l'un des domaines que recouvre l'ingénierie cognitive. L'émergence de cette problématique est récente car, bien que les premiers travaux sur le traitement automatique de la parole soient attribués à A. Graham Bell en 1881 avec le Graphophone (Juang & Rabiner, 2004), l'apparition des premiers systèmes capables d'une reconnaissance vocale automatique date seulement des années 1970 (Juang & Rabiner, 2004; Mc Tear, 2002; Minker & Néel, 2002). Ces recherches étaient issues d'initiatives gouvernementales : un programme américain de la DARPA (projet SUR : « *Speech Understanding Research* ») et, dans les années 1980, le projet japonais d'ordinateur de cinquième génération et le programme européen ESPRIT. Ces programmes de recherche ont permis de mettre au point les méthodes de traitement statistique du signal auditif, nommées *modèles de Markov cachés* (HMM, pour « *Hidden Markov Models* », Levinson, Rabiner & Sondhi, 1983) qui permettent la reconnaissance vocale. A partir de ces initiatives, deux types d'applications commerciales ont vu le jour : (1) les logiciels de dictée vocale (« *voice-activated typewriter* ») d'IBM, Dragon Systems ou Philips, et (2) les services automatiques de télécommunication avec le public développés par les opérateurs téléphoniques (AT&T, NTT DoCoMo, France Télécom, British Telecom, etc.). Les logiciels de dictée vocale consistent principalement en une transcription de la chaîne de parole produite par un locuteur humain en une chaîne de caractères numérisée par l'ordinateur qui exécute le programme. Bien que ces travaux présentent également de l'intérêt, ce chapitre est consacré aux systèmes de télécommunication. L'importance des principes de gestion du dialogue a été progressivement reconnue au cours des années 1990, lors du développement des premiers services (Bernsen, 1994; Brennan & Hulteen, 1995; Bunt, 1994; Whittaker, Brennan & Clark, 1991). Des principes de conceptions ont été établis, d'abord dans le contexte industriel (Maury, 1991; Pontal, 1997), puis par les organismes de normalisation, dont notamment le W3C qui propose et travaille sur des standards pour la navigation vocale et multimodale[1]. Le projet danois Dialogue (Bernsen, Dybkjaer & Dybkjaer, 1998) et le projet européen EAGLES (Gibbon, Moore & Winski, 1997) en sont à l'origine et, entre autres, ont permis d'aboutir au standard de programmation qu'est le VoiceXML[2].

[1] Les activités de ces groupes peuvent être consultées en ligne : http://www.w3.org/Voice/ pour les navigateurs vocaux (« *voice browsers* ») et http://www.w3.org/2002/mmi/ pour l'interaction multimodale.

[2] Cf. http://www.w3.org/TR/voicexml20/

2.2.1 Qu'est-ce qu'un système de dialogue ?

Un système de dialogue est un serveur informatique, généralement distant et accessible via le téléphone ou internet, capable de produire et d'interpréter des énoncés verbaux. Ces énoncés sont vocaux avec un téléphone classique. Ils peuvent être présentés selon d'autres modes sur des terminaux différents. Sur internet, le clavier est utilisé en entrée et l'écran en sortie pour un affichage graphique des réponses. L'intérêt d'un système de dialogue est de délivrer des informations à partir d'une base de données (banques, sociétés de livraisons, etc.) en utilisant le langage courant. La justification principale consiste à dire que ce mode de communication est le plus naturel.

Les premiers systèmes ont été conçus dans un contexte de développement des télécommunications pour être utilisés sur le réseau téléphonique. Dans ce contexte, ils sont désignés sous le nom de *Serveurs Vocaux Interactifs* (SVI, ou encore IVR pour *« Interactive Voice Response »*). Ce sont des serveurs informatiques branchés à un nombre variable de lignes téléphoniques et exécutant un programme de reconnaissance vocale basé sur un vocabulaire constitué d'un nombre définit de mots dans une langue donnée : le modèle de reconnaissance vocale. Ce modèle est appris par le programme à partir d'un échantillon d'enregistrements sonores des mots du vocabulaire. Le programme est contrôlé par un serveur d'applications, capable d'associer des actions à produire à chacun des mots du vocabulaire. La difficulté première réside dans la conception de modèles de reconnaissance vocale fonctionnant en multilocuteur, c'est-à-dire capables de reconnaître n'importe quel individu parlant la langue, dans un vocabulaire de taille acceptable.

Les ressources de calcul des ordinateurs ont, bien sûr, permis d'augmenter progressivement la taille du vocabulaire (jusqu'à 36 000 villes pour la France et plus de 200 000 patronymes pour une version automatique de l'ancien *'12' :* ce numéro permettait, en des temps anciens, de joindre le service des renseignements téléphoniques). Mais cette contrainte a néanmoins imposé de limiter le nombre de mots reconnaissables simultanément, ce qui a influencé le développement des techniques d'interaction dans la mesure où il était d'abord nécessaire d'imposer à l'utilisateur d'employer les termes reconnaissables par le système. Parallèlement, l'impossibilité d'atteindre une reconnaissance parfaite, même dans un vocabulaire restreint, a imposé de développer des techniques d'interaction adaptées (Brennan & Hulteen, 1995). Ainsi, l'optique prise dans la conception se rapportait d'abord à une logique de rattrapage des erreurs (ou "écarts", voir Taleb, 1996). Et ainsi, la conception des systèmes a d'abord consisté à rendre leur utilisation aussi acceptable que possible en tenant compte des limites technologiques du moment. Ces limites ayant été progressivement dépassées, il est devenu possible de chercher à perfectionner le comportement de ces automates. Outre les développements techniques, les propositions sont faites notamment sur la base des modèles psychologiques qui permettent de décrire les processus en jeu dans l'activité de dialogue. C'est la compétence du système qui est interrogée.

A Diversité des systèmes

Plusieurs facteurs favorisent actuellement la généralisation des technologies vocales. Outre le gain en puissance des ordinateurs, qui donne lieu aujourd'hui à une capacité intéressante, l'interconnexion généralisée des systèmes, grâce à *Internet*, permet l'accès à des bases de données diverses, la normalisation des réseaux permet l'accès universel à l'information et la convergence des terminaux permet d'envisager un accès à une information plus riche (Minker & Néel, 2002).

Dès à présent, ces systèmes sont utilisés pour accéder à divers types de contenus disponibles grâce aux réseaux informatiques. Les applications les plus célèbres en langue anglaise sont le système *Pegasus,* qui donne des renseignements sur les horaires d'avion aux Etats-Unis (Glass & Weinstein, 2001), le système *Jupiter*, qui permet d'accéder à des contenus météorologiques (Zue et al., 2000), ou encore *« How May I Help You? »* (*« Comment puis-je vous aider »*), le système de routage d'appels en langage naturel d'AT&T (Gorin, Parker, Sachs, & Wilpon, 1996). Des systèmes équivalents ont également été conçus en français pour des applications du même type (*e.g.* Damnati, Béchet & De Mori, 2007; Sadek, Bretier & Panaget, 1997; Sorin, 1994). Ces systèmes couvrent aujourd'hui des applications dans le domaine des transports (*e.g.* suivi de colis), le domaine bancaire (opérations bancaires et boursières), les renseignements (téléphonie, cinéma, météorologie, etc.), ainsi que d'autres activités de relation clientèle (numéros verts des entreprises, radio/télévision, etc.). En outre, les possibilités multimodales peuvent accroître ce potentiel. Les systèmes de dialogues semblent donc promus à un avenir certain.

B Classification des systèmes

Les systèmes de dialogue ont d'abord été conçus pour communiquer à l'oral. De ce fait, ils sont basés sur un fonctionnement tour à tour comme c'est le cas dans le dialogue humain. On distingue les systèmes à états finis, les systèmes basés sur des gabarits (ou systèmes de formulaires) et les systèmes agents (pour une présentation détaillée, voir Mc Tear, 2002) :

- Dans un ***système à états finis***, ou *« automate à états »*, le dialogue est spécifié sous la forme d'un ensemble d'états de dialogue dont les liens peuvent être représentés par un graphe. Dans chaque état, le système (1) diffuse des *prompts* (fichiers sonores préenregistrés) pour donner les consignes voulues à l'utilisateur, (2) il active tout ou partie du vocabulaire du modèle de reconnaissance vocale en fonction du contenu de la consigne, et (3) il produit une action spécifique pour chaque mot reconnu. L'interaction avec le système consiste à se déplacer d'un état à un autre à l'aide de commandes vocales constituées de mots uniques. La structure est figée, obligeant l'utilisateur à franchir les étapes dans un ordre préétabli. Les prompts diffusés par le système doivent être prudemment rédigés pour s'assembler correctement.
- Dans un ***système à gabarits*** *(i.e.* système de formulaire), les questions posées à l'utilisateur permettent au système de remplir les champs d'un formulaire. De ce fait, la

structure de l'interaction n'est plus prédéterminée et figée, mais dépend de ce que l'utilisateur dit et de ce qu'il omet. Le modèle de reconnaissance vocale est enrichi et inclut une interprétation sémantique, de sorte que la formulation des demandes est libre. Le système fonctionne sur la base de règles de production qui permettent de choisir les actions à produire et les changements d'états. Les états nécessaires à la gestion du dialogue sont plus nombreux, de sorte que la flexibilité de ces systèmes en amplifie également la complexité. Les écarts au dialogue sont toujours traités dans une logique d'erreur, mais les techniques de réparation sont enrichies.

- Les ***systèmes agents*** sont issus de l'Intelligence Artificielle (IA). Ils sont conçus pour permettre une interaction complexe entre l'utilisateur, le système et une application externe. Dans ces systèmes, la communication est vue comme une interaction entre les deux agents (Sadek, 1991; Sadek, Bretier & Panaget, 1997). Le système est doté de croyances, de désirs et d'intentions qui lui donnent la capacité de raisonner sur ses actions, et éventuellement sur celles des autres agents. Un tel système est capable d'utiliser des attentes (*« expectations »*) pour prédire et interpréter les prochains énoncés de l'utilisateur. Ces systèmes permettent également une initiative mixte, permettant à l'utilisateur et au système d'ajouter de nouvelles contributions.

Tableau 2-1 : Classification des systèmes de dialogue

- Types de systèmes	**Systèmes à états finis** *(« finite state-based »)*	**Systèmes à gabarits** *(« template-based »)*	**Systèmes agents** *(« agent-based »)*
- Entrée vocale	Mot-clé ou phrase isolés.	Langage naturel avec détection de concepts.	Langage naturel non restreint.
- Vérification	Confirmation explicite de chaque entrée.	Confirmation explicite ou implicite.	Gestion du processus de « grounding ».
- Modèle du dialogue	Représentation implicite des états d'information dans les états de dialogue. Contrôle du dialogue représenté explicitement dans le diagramme d'états.	Représentation explicite des états d'informations. Contrôle du dialogue représenté par un algorithme de contrôle.	Modélisation des intentions, buts et croyances du système. Histoire du dialogue, prise en compte du contexte.
- Modèle de l'utilisateur	Modèle simple des caractéristiques ou préférences.	Modèle simple des caractéristiques ou préférences.	Modélisation des intentions, buts et croyances de l'utilisateur.

Le Tableau 2-1 (issu de Mc Tear, 2002) propose une description synthétique de ces trois types de systèmes vocaux. Le traitement des *entrées vocales* et la prévention/réparation des écarts (*vérifications*) sont les deux tâches principales accomplies par le système. La modélisation des connaissances (modèles *du dialogue* et *de l'utilisateur* et aussi modèle *de la tâche*, non présent dans le tableau) est également importante. Les problèmes de modélisation sont abordés plus loin, dans la partie dédiée à la conception des systèmes.

C La chaîne de traitement

Différents types de traitement sont nécessaires à la production des réponses du système. Ces traitements sont réalisés les uns à la suite des autres, dans une chaîne modulaire.

Chacun des composants utilise en entrée les informations produites en sortie du composant précédent. Cette architecture linéaire peut cependant fonctionner d'une façon plus subtile dans les systèmes plus évolués. La Figure 2-2 symbolise ces traitements et représente les composantes du système.

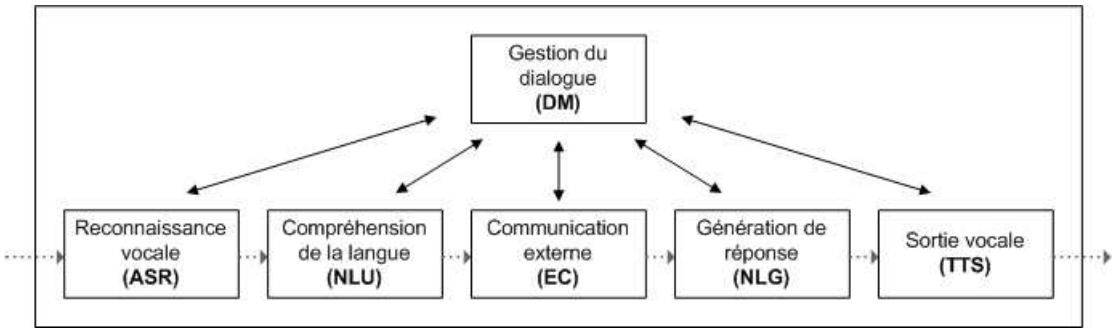


Figure 2-2 : La chaîne de traitement d'un SVI

Les fonctions de ces composantes consistent en une analyse linguistique des demandes de l'utilisateur formulées en langage naturel, qui permet au système d'accéder à des informations, puis de les restituer à l'utilisateur, également sous forme linguistique. Le Tableau 2-2 présente les modules qui composent cette chaîne.

Tableau 2-2 : La chaîne de traitement

Reconnaissance vocale ("Automatic Speech Recognition": ASR)
Le composant de reconnaissance vocale a pour fonction de convertir les énoncés de l'utilisateur, pris dans leur forme auditive (fichiers sons structurés temporellement), en une chaîne de caractères (sous la forme de phonèmes ou de mots).
Compréhension de la langue ("Natural Language Understanding": NLU)
Le composant de compréhension de la langue analyse la chaîne de caractère fournie par la reconnaissance vocale et en dérive le sens de l'énoncé (analyses syntaxique et sémantique). Selon les systèmes, ces analyses peuvent être omises (automates à états) ou confondues, ou l'analyse sémantique peut-être réalisées sans analyse syntaxique.
Gestion du dialogue ("Dialogue Management": DM)
Le composant de gestion du dialogue contrôle le flux du dialogue. Il détermine si suffisamment d'information a été obtenue, communique avec l'application externe et retourne les informations à l'utilisateur. Ce composant est chargé des opérations de contrôle du dialogue et décide des méthodes de prévention et de gestion des erreurs.
Communication externe ("External Communication: EC)
Le composant de communication externe est chargé de communiquer avec la base de données, externe à l'application. Ce composant est chargé de la conversion des demandes de l'utilisateur sous la forme de requêtes d'accès distant.
Génération des réponses ("Natural Language Generation": NLG)
Le composant de génération construit le message de réponse. Il doit (1) sélectionner les informations à inclure, (2) structurer ces informations et (3) et décider de la forme finale (choix des mots et structure syntaxique).
Sortie vocale ("Text-to-speech": TTS)
La production d'une réponse vocale implique de traduire l'énoncé en une chaîne sonore. Les messages peuvent être préenregistrés et concaténés. Mais si le texte est plus variable l'emploi d'une synthèse vocale est requis pour convertir la chaîne de caractères.

D Contrôle du dialogue

La méthode de contrôle du dialogue dépend du fonctionnement de l'ensemble des composantes et, conséquemment, du type de système conçu. Elle régit l'organisation des tours et occupe un rôle fondamental dans le fonctionnement (Mc Tear, 2002). Les deux questions principales sont de savoir qui prend l'initiative dans le dialogue (le *système*, l'*utilisateur* ou une initiative *mixte*) et comment le dialogue est organisé.

Dans un système à états finis, l'initiative revient au système. La structure du dialogue peut être représentée sous la forme d'un réseau d'états de transition simple, voire de structure linéaire. Tout le dialogue est déterminé à l'avance, d'où la logique de traitement des incompréhensions du système comme des « écarts » (Taleb, 1996). Au sens propre, elles écartent le dialogue du *droit chemin*. A l'autre extrémité, les systèmes agents permettent d'entretenir une initiative mixte et d'organiser les échanges en fonction des besoins. On peut dire que la *référence* (l'objet de la transaction) est construite sur un mode plus coopératif (voire collaboratif, *Cf.* Amiel, 2005) car des intentions sont partagées entre l'agent humain et l'agent système. Le dialogue est plus souple.

Il semble évident que les systèmes agents offrent des possibilités plus intéressantes que les autres, mais dans les faits, ils ne sont pas toujours la solution choisie. Tout d'abord, ils imposent une complexité technique supplémentaire puisque l'architecture est plus complexe et puisque des compétences dans un langage de programmation logique (par exemple : « *Prolog* » ou « *LISP* ») sont nécessaires. De plus, une logique de réduction des coûts impose généralement d'aller au plus simple. Les systèmes à états finis et à gabarits sont donc souvent le choix le plus réaliste.

Mais d'autres raisons peuvent justifier le choix d'un système plus simple. Par exemple, Hone et Baber (1995) ont observé l'effet du mode de contrôle sur la durée du dialogue. Bien qu'ils aient enregistré des dialogues plus longs dans une interaction contrôlée par des questions oui/non successives, ils ont également trouvé des dialogues longs, dans le cas de l'interaction mixte, quand le nombre d'erreurs de reconnaissance était important (cette condition était manipulée dans cette expérience). Dans une autre expérience (Potjer, Russel, Boves, & Os, 1996), dans laquelle la reconnaissance vocale n'était pas simulée, la version d'initiative système fonctionnait en reconnaissance de mots isolés alors que la version d'initiative mixte fonctionnait en reconnaissance de parole continue. Dans cette expérience, le nombre de tours de parole était plus bas dans la version avec initiative mixte car les utilisateurs formulaient des requêtes plus complètes, mais les erreurs de reconnaissance vocale étaient également plus fréquentes dans cette version, de sorte que les dialogues n'étaient pas plus courts. De plus, les utilisateurs étaient satisfaits du système dans les deux modes d'interaction. Ainsi, des contraintes apparaissent quel que soit le type de système utilisé, mais les buts sont atteignables dans tous les cas.

Les solutions offertes dans un *dialogue dirigé par le système* permettent également de bien guider l'utilisateur, ce qui est d'autant plus utile quand il est novice. Ainsi, ce mode de

contrôle est intéressant pour des tâches simples et bien structurées. C'est pourquoi des systèmes à états finis et à gabarits ont été créés pour des tâches telles que la recherche dans un répertoire, la gestion de calendrier, pour des questionnaires et pour la recherche d'informations. Quand les tâches se complexifient, par exemple si elles impliquent pour le système d'acquérir plusieurs critères de recherche, les systèmes à gabarits permettent d'assouplir l'ordre des questions et de rétablir une sensation de dialogue pour l'utilisateur. Cependant, ces systèmes sont difficiles à concevoir si les tâches à réaliser sont sources de variabilité, tant du point de vue de l'utilisateur et du rôle de ses connaissances, que du point de vue de la tâche en elle-même et de ses implications. Par exemple, la prise de rendez-vous avec les médecins d'un hôpital peut parfois imposer des connaissances médicales dont la prise en compte par un automate n'est pas aisée. Ces activités, telles que la négociation et la planification ne sont pas permises avec des automates simples. Dans ce cas, l'approche agent peut permettre de formaliser ces tâches complexes. Mais quel que soit le type de tâche développé et le système utilisé pour la conception, la méthodologie occupe une place importante car elle permet aux différents acteurs du projet de coordonner leurs actions pour stabiliser le comportement du système. Ce point sera abordé après avoir présenté l'approche agent.

2.2.2 Les systèmes agent

L'approche agent se focalise sur la modélisation du processus de dialogue en tant que collaboration entre plusieurs agents. Il existe différentes façon d'aborder ce problème, telles que la modélisation de l'activité en tant qu'activité de recherche de preuves de théorèmes (*« theorem proving »*), la modélisation par plans et les agents conversationnels. Les approches les plus abouties se trouvent dans la catégorie des agents conversationnels qui se base sur une modélisation des *états mentaux* des interlocuteurs. Quel que soit le type de modèle utilisé, l'objectif est de représenter l'activité sous la forme d'étapes discrètes et de règles de transformation qui peuvent opérer à chacune de ces étapes.

A Le modèle BDI

L'approche la plus influente pour la conception des agents conversationnels se trouve ses fondements dans l'architecture BDI (*« Believe, Desire, Intention »*, ou en français, *« Croyance, Désir, Intention »*), proposée par Bratman, Israel et Pollack (1988), qui permet de modéliser les actions conversationnelles (*Cf.* Figure 2-3, tirée de Allen, 1995). A partir de ses croyances et de ses obligations, l'agent sélectionne des buts communicatifs. Il décide quels actes réaliser et génère l'énoncé correspondant. Il analyse la réponse du manager humain du dialogue et met à jour ses croyances sur l'état du discours et sur ses obligations.

Ce modèle est une formalisation logique de la boucle fonctionnelle (Cf. Gibson, 1979; Rasmussen, Pejtersen & Goodstein, 1994) qui opère dans l'activité de dialogue et dont l'unité est la paire adjacente *énoncé-réponse*. Il offre un cadre conceptuel pour la conception

d'agents capables de considérations rationnelles au sujet de leurs actions à partir, à la fois, de leurs *obligations*, de leurs *connaissances* et de leurs *buts*.

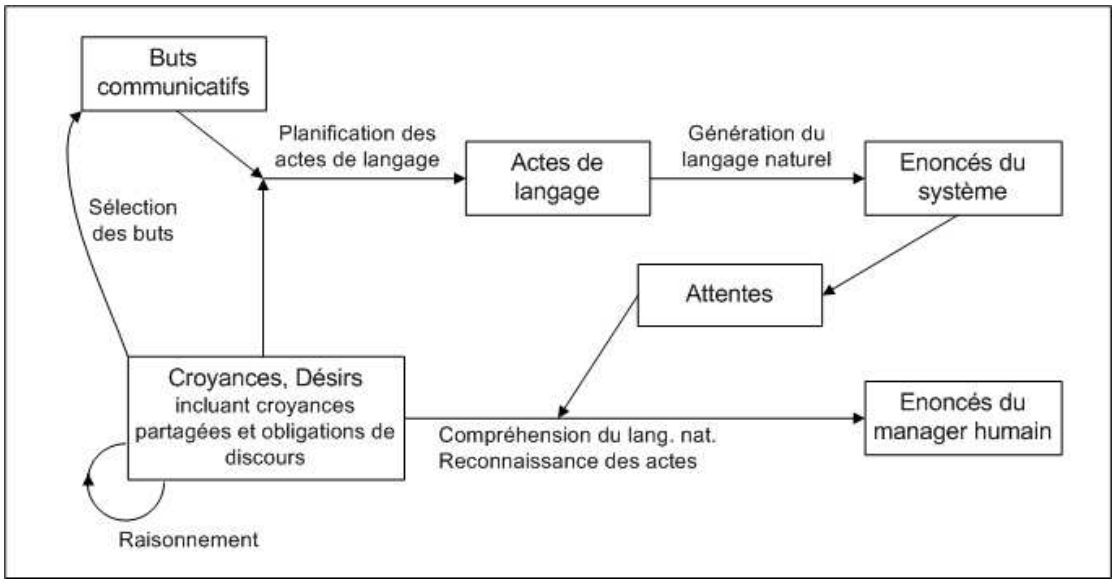


Figure 2-3 : Le modèle BDI (*« Believe, Desire, Intention »*)[1]

B Théorie de l'interaction rationnelle

La *théorie de l'interaction rationnelle* (Sadek, 1991; Sadek, Bretier & Panaget, 1997) utilise ces concepts pour formaliser l'activité de dialogue (voir aussi, Bretier, 1995; Panaget, 1996). L'objectif principal est de donner aux systèmes dialoguant la capacité de produire des comportements coopératifs, tels que celui qui consiste à fournir une réponse suggestive. Par exemple : *« Non, il n'y a pas de serveurs d'emploi pour Calais. Par contre, il y a un serveur d'emploi pour le Pas-de-Calais et un serveur d'emploi pour Lille. L'un d'eux vous convient-il ? »* (Sadek, Bretier & Panaget, 1997). La formalisation est basée sur une analyse des principes d'action et de coopération développée par Cohen et Levesque (1990) et étendue par Sadek (1991).

Pour être capable de produire ce comportement, le système doit accomplir une séquence d'actions lors de chaque demande de l'utilisateur :

- Il infère que l'intention de l'utilisateur est de connaître la réponse à sa demande ;
- Il adopte l'intention que l'utilisateur puisse éventuellement connaître la réponse à sa demande ;
- Il adopte l'intention d'informer l'utilisateur de la réponse à sa demande.

L'intérêt de faire adopter ces trois intentions à l'agent est de lui permettre de faire des choix quand les objectifs sont atteints ou quand, au contraire, ils échouent. Les intentions

[1] La figure proposée est une reproduction exacte (hors traduction en français) de la figure proposée par Bratman, Israel ou Pollack (1988). Le terme « intention » n'y apparaît pas directement. La direction de la flèche du bas pourrait être inversée.

permettent de représenter les causes de l'action dans l'environnement. Généralement, l'intention est notée « $I_i\Phi$ » (*« l'agent i a l'intention que Φ »*). On distingue trois notions fondamentales (Tableau 2-3).

Tableau 2-3 : Les trois types d'intention dans la théorie de l'interaction rationnelle

Le ***But à réaliser***	C'est une propriété que l'agent croit non réalisée actuellement et qu'il souhaite voir se réaliser dans le futur. On note $AG_i\Phi$ *(« AG »,* pour *« agent goal »*)
Le ***But persistant***	C'est un but à réaliser que l'agent abandonnera lorsqu'il croira qu'il est réalisé ou impossible à réaliser On note $PG_i\Phi$ *(« PG »,* pour *« persistent goal »*)
L'***Engagement coopératif***	Un agent souhaite accomplir le début de toute suite d'actions (éventuellement multi-agent) qui peut aboutir à la satisfaction d'une propriété jusqu'à ce qu'il pense cette propriété réalisée ou impossible à réaliser. Cette propriété implique qu'un agent peut s'engager dans une série d'actions jusqu'à ce qu'il pense qu'une certaine propriété est satisfaite, auquel cas il peut abandonner ce but. Cette proposition de Sadek (1991) est une reformulation de la notion introduite par Cohen et Levesque (Cohen & Levesque, 1990).

Le *modèle de l'action* établit les correspondances entre l'état mental de l'agent et les évènements du monde. Il permet de représenter les évolutions de l'environnement. Il incorpore une dimension temporelle puisqu'il réfère à l'exécution de l'acte. Les évènements ne sont pas datés sur un axe continu mais relativement les uns aux autres (passé, présent, futur). Il définit les pré-conditions de faisabilité (PF) et l'effet rationnel (ER), l'agent auteur de l'acte (i), le destinataire (j) et le contenu (Φ). Le modèle est défini comme suit :

<i,Acte(j,Φ)>
PF : PF(a)
ER : ER(a)

Les actes définis dans Artimis consistent à informer (INFORME, CONFIRME, INFORMESI, INFORMEREF) et à demander (DEMANDE). Une présentation détaillée peut être trouvée dans une thèse récente (Baudoin, 2007), en français. Dans chacun de ces actes, les fonctions PF(a) et ER(a) sont définies sur la base des connaissances (B pour « Believe ») de l'agent i, auteur des actes ($B_i\Phi$), et de l'agent j, destinataire des actes ($B_j\Phi$). On distingue les *effets directs* des *effets indirects* de l'acte. Les effets directs sont les conséquences de l'action visées par l'agent. Les conséquences indirectes de cette action, ou « effets de bord », ne sont pas prises en compte dans la théorie. En pratique, cela implique que les effets indirects ne font pas partie des connaissances de l'agent et qu'il ne lui est pas possible de les prendre en considération dans ses raisonnements.

Le comportement de l'agent est basé sur un jeu d'*axiomes* qui formalisent les principes de rationalité et de coopération. Ils établissent les rapports entre les attitudes mentales de l'agent. Le Tableau 2-4 présente les principaux axiomes utilisés dans Artimis (voir, Baudoin, 2007, pour une présentation ; Louis, 2002, pour une revue complète). Il s'agit d'une révision des principes par Sadek et al. (1997), établie sur la base de divers retours d'expérience. Le premier principe repris dans le tableau exprime la *consistance des croyances*.

Tableau 2-4 : Formalisation des axiomes de rationalité et de coopération

Principes de rationalité	
$I_i\Phi \Rightarrow B_i\neg\Phi$	*L'agent i ne peut avoir l'intention de réaliser un but que s'il croit que ce but n'est pas encore réalisé.*
$I_i\Phi \Rightarrow I_i\text{Fait}(a_1\|\ldots\|a_n)$ Où $\Phi = ER(a_k)$	*Si l'agent i a comme connaissance que l'une de ses actions peut le conduire à l'un de ses buts, alors il sélectionnera cette action.*
$I_i\text{Fait}(a) \Rightarrow (B_i\text{Faisable}(a) \vee I_iB_i\text{Faisable}(a))$	*L'agent i cherche à satisfaire les pré-conditions d'un acte quand il sélectionne cet acte.*
Principes de coopération	
$B_i((PG_j\Phi \wedge \neg PG_i\neg\Phi) \Rightarrow I_iB_i\Phi)$	*L'agent i tente de réaliser les buts de son interlocuteur s'il en a conscience.*
$B_jI_iB_i\Phi(i) \wedge \neg B_j\neg\Phi(i) \Rightarrow B_j\Phi(i)$	*L'agent i est amené à croire une proposition par le fait de penser qu'un autre agent la croit.*
$B_i((\Phi \wedge B_j\neg\Phi) \Rightarrow I_iB_i\neg\Phi)$	*L'agent i corrige ses croyances quand il constate une incompatibilité avec les faits.*
$\Rightarrow$ = implication ; $\wedge$ = et ; $\vee$ = ou ; $\neg$ = négation logique ; | = alternative	

Le fonctionnement de l'agent est basé sur un diagramme d'états-transition cyclique proche du modèle BDI (Bratman, Israel & Pollack, 1988) présenté dans la Figure 2-3. Il suppose la délivrance d'un message à l'utilisateur, l'attente d'une réponse, son acquisition et son interprétation, une phase de raisonnement incluant les actions applicatives et une phase de contrôle déterminant la forme finale de la réponse apportée par l'agent.

C Avantages, limites et perspectives

La souplesse de fonctionnement du système provient du fait que, dans cette approche, la structure du dialogue n'est pas écrite à l'avance. Comme le remarque Mc Tear (2002), ici la structure émerge de façon dynamique comme une conséquence des principes rationnels de l'interaction coopérative. Ainsi, cette théorie fournit les bases d'une approche explicative du dialogue, en termes de plans, de buts et d'intentions. Même les plans utilisés ne sont pas des schémas d'action prédéterminés. Il est, par exemple, possible de tenir compte de la volonté de l'utilisateur de respecter la confidentialité d'une information. Pour agir, l'agent respecte un *équilibre rationnel* entre ses propres attitudes mentales et celles des autres agents.

D'un autre coté, au-delà de la complexité engendrée par le développement d'un tel système, des efforts de recherche sont encore nécessaires du point de vue de l'intégration du contrôle du dialogue avec les connaissances du domaine et de la tâche (Mc Tear, 2002). Divers travaux ont été engagés dans ce sens de façon à optimiser, par exemple, la reconnaissance de plan (Louis, 2002) puis la réflexion sur la notion de capacité (Devooght, 2007; G. Meyer, 2006), l'acquisition de connaissances sémantiques (Duclaye, 2003), l'apprentissage et l'optimisation de séquences d'actions (Baudoin, 2007) ou encore le traitement des références (Saget & Guyomard, 2007). Des travaux de formalisation de la notion d'attente ("expectation", voir par exemple Castelfranchi, 2007; Jokinen & Hurtig, 2006) présentent également

beaucoup d'intérêt. L'évolution va dans le sens de la diversification des capacités et permet de représenter différents processus du raisonnement abstrait.

Même si un système de dialogue n'intègre pas ces principes dans son fonctionnement en temps réel (s'il est conçu comme un automate) la représentation formelle des états mentaux conserve une pertinence pour la représentation conceptuelle des dialogues, lors de l'analyse préalable et des phases de conception du projet, et plus encore dans une démarche de recherche.

2.3 Problèmes de conception

La conception d'un système de dialogue est un processus qui intègre différents acteurs. Ceux-ci doivent coordonner leur travail et vérifier leurs hypothèses dans des cycles de développement au cours desquels ils intègrent les contraintes du projet et utilisent les connaissances à leur disposition, sous la forme de modèles prédictifs, qui leur permettent de définir le fonctionnement final du système.

Cette partie va permettre de revenir sur cette démarche de conception, les principes et les modèles de référence en psychologie, linguistique et intelligence artificielle sur lesquels elle se base pour proposer les systèmes de dialogue tels que nous les connaissons.

2.3.1 Cycle de développement : spécification, conception, évaluation

Le développement d'un système de dialogue est un cas particulier des méthodes d'ingénierie du logiciel, qui intègre certaines méthodes et certains critères d'évaluations spécifiques (magicien d'Oz, modélisation de la langue, etc.). Les méthodes de spécification, de conception, de développement et d'évaluation font l'objet de standards développés dans les projets Dialogue et EAGLES (Bernsen, Dybkjaer & Dybkjaer, 1998; Gibbon, Moore & Winski, 1997) et qui peuvent servir de référence aux praticiens. Il faut cependant noter que l'existence de standards n'implique pas une méthodologie standardisée. Par exemple, Allemandou (2007) note qu'il n'existe pas de consensus en ce qui concerne le paradigme d'évaluation entre approches objectives et subjectives (problème pour lequel il propose des solutions).

Le développement d'un système de dialogue implique de décider des tâches que le système devra réaliser, de spécifier la (ou une) structure de dialogue correspondante, de déterminer le vocabulaire à reconnaître et de le structurer en un modèle de langage, puis de développer informatiquement et d'implémenter la solution. Les énoncés du système sont conçus au cours de la phase de spécification du dialogue, parfois sur la base de simulations ou de Magiciens d'Oz, et plus souvent à l'aide d'organigrammes (*« Flowcharts »*), ou parfois de tableaux, décrivant la structure. Ils sont rédigés par des ergonomes (*« designers »*) qui veillent à la fluidité du dialogue, la cohérence, etc. pour faciliter l'utilisation des fonctions du système aux

utilisateurs. Le cahier des charges fournit par le maître d'œuvre du projet (l'entreprise qui souhaite fournir le service à ses clients) indique les fonctionnalités à développer et les contraintes à prendre en compte. Le dialogue et les énoncés conçus sont ensuite révisés dans les phases d'évaluation qui accompagnent l'implémentation, puis suite aux tests utilisateurs et également au cours de la vie du service. La conception d'un système agent plutôt qu'un automate ne modifie pas le processus de conception mais implique simplement une complexité plus grande. Par exemple, dans le cas du système agent, le nombre d'énoncés différents est plus important et diverses incises sont possibles, ce qui permet au système de couvrir un champ comportemental plus large.

Pour la spécification, le projet Dialogue (Bernsen, Dybkjaer & Dybkjaer, 1998) prévoit l'utilisation de différents documents regroupés en deux classes pour représenter, d'une part, la structure et les contraintes (DSD pour *« Design Space Development »*), et d'autre part, les raisonnements spécifiques à chaque problème posé (DR pour *« Design Rationale »*). De même, EAGLES (Gibbon, Moore & Winski, 1997) recommande une description formelle et explicite des dialogues proposés. L'objectif de ces documents est d'expliciter les choix faits par l'équipe au cours du processus de conception. Il s'agit en réalité, souvent implicitement (voir, par exemple, Karsenty, 2000), de réaliser une analyse fonctionnelle de façon à prendre en considération, autant que possible, l'ensemble des facteurs qui interviendront au cours de l'utilisation ultérieure du système par le public cible.

2.3.2 Principes de conception et de modélisation pour le DHM

Deux articles essentiels permettent de préciser les contraintes à prendre en compte (Clark & Brennan, 1991) et les connaissances à modéliser dans les systèmes (Brennan & Hulteen, 1995) pour une gestion efficace de l'interaction en dialogue vocal. Mais l'introduction de la multimodalité diversifie les contraintes en jeu.

A 'Grounding' en communication

L'article de Clark et Brennan (1991) identifie les différentes sources de coût (pour les utilisateurs) dans les communications et propose une explication de la gestion de l'effort basée sur la notion de *moindre effort collaboratif* (Clark & Wilkes-Gibbs, 1986).

Clark et Brennan (1991) fournissent des exemples prélevés dans les communications médiatisées (email, téléconférence, etc.) et dans le DHM. Après avoir rappelé les principes théoriques liés au *processus de grounding* (*Cf.* chapitre 1), les auteurs présentent les contraintes qui peuvent jouer sur ce processus (*Cf.* Tableau 2-5). Pour les auteurs, chaque média associe ces contraintes d'une façon spécifiques. Ils illustrent ce fait dans sept médias (*« medium »*) différents : (1) *face à face*, (2) *téléphone*, (3) *vidéoconférence*, (4) *téléconférence*, (5) *systèmes d'information sur ordinateur*, (6) *email*, (7) *lettres*. Dans le cas du téléphone et du DHM, il y a : *audibilité*, *co-temporalité*, *simultanéité* et *séquentialité*.

Tableau 2-5 : Contraintes jouant sur le processus de *'grounding'*

Coprésence	Les interlocuteurs partagent-ils le même environnement ?
Visibilité	Les interlocuteurs sont-ils visibles l'un pour l'autre ?
Audibilité	Les interlocuteurs peuvent-ils communiquer par la parole ?
Co-temporalité	La transmission des messages est-elle immédiate ou différée ?
Simultanéité	Les interlocuteurs peuvent-ils transmettre des messages en même temps ou seulement alternativement ?
Séquentialité	Les séquences d'échanges peuvent-elles être interrompues, par exemple, par des communications avec d'autres partenaires ? Quel est le rythme des échanges ?
Reconsultabilité	Le destinataire d'un message peut-il revoir ce message par la suite ?
Révisabilité	Le producteur d'un message peut-il réviser le contenu de son message avant la validation ou l'envoi ?

Les auteurs expliquent que l'absence de ces caractéristiques engendre un coût pour les interlocuteurs parce qu'ils doivent trouver des techniques de contournement. Onze types de coûts sont envisagés : (1) coûts de *formulation*, (2) de *production*, (3) de *réception*, (4) de *compréhension*, (5) de *démarrage*, (6) de *délai*, (7) d'*asynchronie*, (8) de *changement de locuteur*, (9) d'*affichage*, (10) de *faute* et (11) de *réparation*.

Le *principe de moindre effort collaboratif* intervient dans la gestion de ces sources de coût. Pour les auteurs, chaque source de coût entretient une relation particulière avec les niveaux de coordination (*Cf.* chapitre 1, Tableau 1-8). Par exemple, lors d'une communication clavier, les utilisateurs tendent à fournir des explications complètes sans vérifier que l'interlocuteur est bien à l'écoute ou comprend, ou valide, chaque étape. Au téléphone, au contraire, les confirmations et les signes d'attention sont très réguliers. Ainsi, chaque type de coût exerce des contraintes sur les coordinations entre les partenaires au sein du processus collaboratif.

Finalement, les auteurs précisent qu'il existe une relation (*« interaction »*) entre le *média* utilisé et le *but* à atteindre. Par exemple, la communication en face-à-face est préférée pour des activités de négociation et de recherche de consensus, alors que pour des tâches de planification, de programmation d'horaires (*« scheduling »*) ou de rédaction de rapports, c'est l'email qui est préféré. Mais les auteurs précisent que cette interaction entre *média* et *but* dépend de la forme que prend le processus de grounding dans l'interaction.

B 'Grounding' en DHM

Brennan et Hulteen (1995) utilisent les notions présentées dans l'article précédent pour proposer un '*modèle du processus de grounding avec un système de DHM*'. Ils étendent d'abord la liste des niveaux de coordination (*Cf.* chapitre 1. Tableau 1-8) et considèrent huit états différents :

Tableau 2-6 : Les huit niveaux de coordination, d'après Brennan et Hulteen (1995)

(0)	*'Absence d'activité'*	Pas de détection de parole de l'utilisateur ;
(1)	*'Détection d'activité'*	Détection d'une parole ;

(2)	*'Ecoute'*	Identification des mots ;
(3)	*'Analyse grammaticale'*	Structuration de l'énoncé de l'utilisateur ;
(4)	*'Interprétation'*	Analyse sémantique de l'énoncé de l'utilisateur ;
(5)	*'Intention'*	Conversion de l'énoncé en une commande dans le système ;
(6)	*'Action'*	Exécution de la commande ;
(7)	*'Rapport d'action'*	Confirmation de l'exécution de la commande.

Pour chacun de ces états, des *réponses* (*« feedback »*) du système sont prédéfinies, sous la forme d'énoncés standardisés, suffisants pour expliquer à l'utilisateur l'état de prise en compte de sa demande, soit l'état de compréhension actuel. Les auteurs établissent que le système ne donne que le niveau le plus avancé. Par exemple, s'il est dans l'état (3), il diffuse uniquement l'énoncé correspondant à cet état, et non la chaîne (0)-(1)-(2)-(3).

- ***Modèle de la tâche***

En parallèle, un *'modèle de la tâche'* associe le comportement le plus adapté en fonction du risque d'erreur. Si une tâche est critique, un niveau de feedback supérieur à celui dans lequel se trouve le système pourra être choisit. Par exemple, si l'utilisateur demande l'appel de Joseph, qui est dans son répertoire, et s'il y a plusieurs Joseph dans ce répertoire, le système pourra choisir le Joseph qui est appelé le plus fréquemment. Mais il peut être définit qu'entre minuit et huit heure du matin, un niveau de confirmation supplémentaire est désirable.

- ***Modèle de l'utilisateur et historique du dialogue***

Enfin, un *'modèle de l'utilisateur'* est défini sur la base de *'*l'*historique du dialogue'*. L'historique décrit la forme de surface des commandes de l'utilisateur et l'intention qui en a été déduite. Un *'critère de grounding'* permet d'adapter le niveau du feedback fournit par des incrémentations successives. Si ce critère est bas, le niveau de feedback tend vers (0) ; s'il est haut, le critère tend vers (7). Par défaut, le critère est définit à (3) au début du dialogue avec un utilisateur inconnu. Au fur et à mesure des échanges, si l'utilisateur accepte les réponses du système, le critère de grounding baisse. Inversement, si l'utilisateur corrige le système, le critère de grounding monte. Cela permet de définir un comportement auto-adaptif du système indépendamment de la tâche.

Les auteurs concluent que ce modèle permet une conception systématique des énoncés du système basée sur une modélisation de *l'utilisateur*, de *la tâche* et de *l'historique du dialogue*. L'avantage de cette approche est d'être généralisable. Les auteurs précisent également qu'en DHM ce type de *'feedback'* adaptatif est indispensable à une interaction de qualité.

Dans cette optique, d'autres études sont menées en intelligence artificielle. C'est le cas par exemple des travaux de Walker et al. (2004) qui se focalisent particulièrement sur le *modèle de l'utilisateur* pour adapter automatiquement le contenu des énoncés du système aux utilisateurs. Ce modèle permet de dépasser le *modèle du langage* (d'ordre linguistique) nécessairement présent dans un système de DHM et de s'intéresser aux *'préférences'* de chaque utilisateur. Ces préférences peuvent porter sur le type de critère qui importe à

différentes personnes (prix ou type de nourriture pour des restaurants, par exemple) ou sur la valeur de ces critères (nourriture chinoise ou française).

C Conclusion

Deux types de confusion apparaissent dans les notions présentées. Tout d'abord, entre les notions d'*effort* et de *coût*, qui font également intervenir la notion d'*affordance*. Et par ailleurs, entre la notion de *modèle de l'utilisateur* et celle de *préférence*.

Le principe de *'moindre effort collaboratif'* montre que les interlocuteurs cherchent à optimiser les actions qu'ils produisent l'un envers l'autre dans le but d'accéder, à moindre frais, à une compréhension commune (soit à un « terrain commun »). Bien que ce principe renvoie à un aspect quantitatif (la quantification de l'effort), la notion sous-jacente est bien l'*action coordonnée* des interlocuteurs, qui peut également être décrite d'un point de vue qualitatif (en fonction de ses buts et de ses effets). En revanche, la notion de 'coût' limite les propriétés de l'action à l'aspect quantitatif. Cela incite, si l'on n'y prend pas garde, à ne considérer que cet aspect de coût lors de la modélisation de l'activité, ce qui est une réduction de sens. Cette réduction devient problématique quand d'autres auteurs citent cet article, car il est réputé décrire les « affordances » des médias de communication (voir par exemple, Whittaker, 2003a). Il est vrai que la problématique des affordances dans les communications médiatisées renvoie aux notions abordées par Clark et Brennan (1991), mais le terme « affordance » n'est pas utilisé dans leur article. Les tâches (*« purposes »*) évoquées par les auteurs (envoi d'email, négociation, etc.) sont des tâches qui relèvent d'un niveau de granularité large, qui tolèrent une description généraliste du coût qu'elles nécessitent. La notion d'affordance renvoie à des processus plus fins, tels que, par exemple, la relation main-clavier au court de l'envoi d'un email. Dans ce cas, il est nécessaire de décrire l'action plus en détails. Ces notions relèvent de points de vue différents et la méthodologie expérimentale mise en place pour l'observation doit en tenir compte. La notion de coût est abordée au chapitre 4 sous l'angle de la *charge cognitive* de l'utilisateur.

Au sujet de la notion de *préférence*, le *modèle de l'utilisateur* proposé par Brennan et Hulteen (1995) ne portait pas tant sur les préférences des utilisateurs que sur leur performance dans la tâche. L'avantage est de se baser sur des observations comportementales réelles de façon à reconnaître les patterns comportementaux privilégiés d'un utilisateur. Or, la conversion en une notion de *préférence* a le désavantage majeur de résumer ces patterns à des valeurs figées dans une base de données. Par exemple, le choix d'une couleur pour un objet ne veut pas dire que cette couleur devra toujours être associée à cet utilisateur. Contrairement à ce principe, des travaux sur la modélisation de l'utilisateur à partir de son niveau de familiarisation avec le système ont porté sur le type de formulation adopté par l'utilisateur dès le premier tour de parole. Les utilisateurs experts d'un service tendent à formuler des requêtes plus courtes (Bretier, Le Bigot, Panaget, & Sadek, 2004). Ces travaux permettent de lier le *modèle de l'utilisateur* au *modèle de la langue*. Cette approche offre des perspectives fonctionnelles plus intéressantes que la notion de préférence.

2.3.3 La conception des énoncés du système

Les travaux qui ont été présentés relèvent du fonctionnement général du système. Ils permettent de structurer les connaissances mais ne définissent pas la manière dont le système doit agir auprès de l'utilisateur. Cette partie permet d'aborder les principes de conception des énoncés adoptés spécifiquement dans le cadre du développement des systèmes de dialogue.

A Décomposition des actes du système

Dès le début des années 1980, Nievergelt et Weydert (1980) faisaient remarquer qu'il est préférable de s'intéresser aux problèmes généraux soulevés par la structure des commandes plutôt que d'envisager des collections de fonctions spécifiques à différents contextes.

Pour Nievergelt et Weydert (1980), les réponses nécessaires aux utilisateurs d'un système interactif répondent aux questions : Où suis-je ? Que puis-je faire ici ? Comment suis-je venu ici ? Où puis-je aller ? Pour donner ces réponses en permanence aux utilisateurs, les auteurs proposent trois concepts :

- Les ***sites*** sont les espaces de données auxquels l'utilisateur accède. Un site est un sous-ensemble des données du système, disponibles à un moment donné ;
- Les ***modes*** sont les fonctions actives dans cet état. Les modes disponibles à un moment donné sont un sous-ensemble des fonctions disponibles via le système ;
- Les ***traces*** (*« trails »*) correspondent aux séquences passées de *modes* utilisés dans les différents *sites* visités.

Les auteurs indiquent que les aides occupent également un statut particulier dans ce système puisqu'elles permettent à l'utilisateur de connaître les différents modes disponibles dans un certain site. A partir de ce système, les auteurs montrent qu'il est possible de concevoir une interface structurant ses ressources correctement et de façon systématique. On peut remarquer que ce système décrit les *'évènements passés'*, les *'données présentes'* et les *'actions futures'*. Il catégorise les données du système et les intègre dans le processus d'interaction que l'utilisateur entretient avec elles. Ainsi, il offre une lecture simple du processus en jeu, qui permet d'identifier rapidement les principales fonctions des éléments qui constituent l'énoncé. Bien que cette catégorisation soit peu connue, elle était mentionnée dans un article *princeps* de Lewis et Norman (1986) comme un formalisme intéressant pour le développement d'interfaces veillant à la prévention et à la correction des erreurs ; préoccupation qui s'avère cruciale en dialogue.

Cette catégorisation des éléments à présenter à l'utilisateur est connue des concepteurs de systèmes de dialogue. Dans les équipes de France Télécom R&D, les termes utilisés sont : (1) *« feedback »* pour les *'trails'*, (2) *« réponses »* pour les *'sites'* et (3) *« relances »* pour les *'modes'*. Ces notions ont été appliquées pour la conception du matériel qui sera présenté dans la partie expérimentale. Pour éviter des confusions liées à la polysémie du terme

'feedback' les termes retenus sont : ***« écho »***, ***« réponse »*** et ***« relance »***. Le terme *'écho'* est issu de l'article de Brennan et Hulteen (1995, p. 144).

B Décomposition des fonctions dialogiques

Outre cette décomposition structurelle des actes qui composent l'énoncé, des analyses plus complètes des fonctions qu'il remplit ont été proposées, notamment par Allwood et ses collaborateurs (Allwood, 1995; Allwood, Nivre & Ahlsén, 1992; Allwood, Traum & Jokinen, 2000; Black et al., 1991). Une synthèse de ces analyses, issue de travaux collaboratifs entre les auteurs, est proposée par Bunt (1994) qui se fixe l'objectif d'un contrôle du dialogue par les systèmes automatiques.

Pour Bunt (1994), le dialogue est une activité pendant laquelle chacun des participants réalise deux tâches parallèlement (Cf. Figure 2-4) :

- La première est la *'tâche sous-jacente'* (*« underlying task »*). Elle correspond au but de la conversation, par exemple, obtenir un billet de train ;
- La seconde est la *'tâche communicative'* (*« communicative task »*). Elle consiste à veiller à tous les aspects qui requièrent une attention constante.

Ces deux sous-tâches sont entremêlées et doivent être gérées en permanence au cours de l'interaction. De plus, chaque *'unité d'information'* produit des effets relativement à plusieurs de ces fonctions.

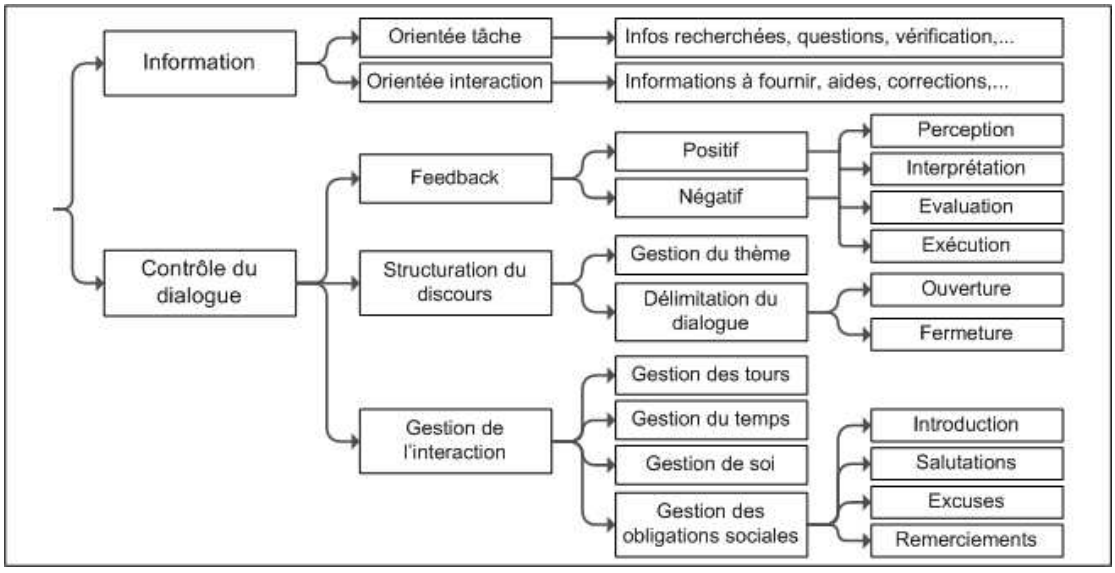


Figure 2-4 : Les fonctions de contrôle du dialogue[1]

Bunt précise qu'un point important et problématique est celui de la *complétude* de l'analyse. Le *'système de fonctions de contrôle'* présenté dans la Figure 2-4 n'est qu'une simplification adaptée aux tâches de recherche d'information qui n'impliquent que deux interlocuteurs. Pour des tâches plus complexes et des groupes sociaux plus importants, ce système est adaptable

[1] Pour une version plus complète, voir : http://let.uvt.nl/general/people/bunt/docs/dit-schema2.html

(*« customizable »*). L'aspect multifonctionnel des énoncés implique une complexité importante qui ne doit pas être ignorée. La conception de systèmes de dialogue utilisant la « langage naturel » suppose une gestion des fonctions naturelles du dialogue.

2.4 Intégration de la multimodalité dans le DHM

La possibilité d'intégrer la multimodalité dans des systèmes vocaux est une perspective récente liée à la convergence des réseaux et au développement des terminaux de taille réduite. L'apparition de cette perspective renouvelle les questionnements précédents sur les connaissances qui doivent être formalisées et implémentées dans les systèmes.

Dans le cadre de la théorie de l'interaction rationnelle, la thèse de Clémente (2004) a permis de définir les pré-conditions de faisabilité (« PF ») des actes du système lorsqu'il utilise plusieurs canaux pour présenter l'information. Ce travail définit l'environnement matériel de l'agent et lui permet de connaître, dans l'action, les moyens à sa disposition. Mais les effets rationnels (« ER ») des actes multimodaux ne sont pas définis dans ce travail, car ils doivent d'abord être mis en évidence. Dans cette perspective, d'autres travaux ont été consacrés au processus de génération des réponses et à la conception des stratégies de dialogue (Horchani, 2007; Horchani, Fréard, Jamet, Nigay, & Panaget, 2007b). Ils ont permis de réfléchir à la conception d'un outil qui pourrait servir à tester les effets des différentes stratégies du système sur la performance et la collaboration. *La thèse présente s'inscrit en continuité directe avec ces travaux*.

Par ailleurs, plusieurs équipes travaillent actuellement au développement des systèmes de dialogue dans la perspective d'une formalisation des fonctions de l'énoncé pour améliorer le processus de génération des réponses (*e.g.* Keizer & Bunt, 2006, 2007a; Sun et al., 2008). Ces études doivent être différenciées des travaux sur le développement des *agents conversationnels animés* (ACA). Il existe des liens conceptuels entre ces travaux, mais les contextes techniques diffèrent et imposent de les distinguer. Les travaux sur les ACA peuvent également porter sur les modalités de la communication ; mais dans ce contexte, c'est surtout la reproduction des principes de communication homme-homme dans le cadre de la coopération homme-machine qui est visée. Cette problématique est liée au domaine des CMC (*« computer mediated cooperation »*) dont le but est la communication *'au travers'* des technologies plutôt que la communication *'avec'* les technologies (Baker, 2003). L'intérêt se porte sur l'utilisation des gestes, (*e.g.* Buisine & Martin, 2007; Cassell, Kopp, Tepper, Ferriman, & Striegnitz, 2007; Lefebvre, 2008) et sur l'expression des émotions (*e.g.* Martin, Niewiadomski, Devillers, Buisine, & Pelachaud, 2006; Ochs, Niewiadomski, Pelachaud, & Sadek, 2006). Il s'agit ici d'ajouter de nouveaux modes d'expression ou de nouveaux comportements à l'agent, alors qu'il s'agissait dans le cas précédent (Keizer & Bunt, 2006) d'optimiser la gestion du processus dialogique. Mais cette distinction ne saurait être strictement exclusive. Les perspectives sont communes.

Différentes approches existent au sein même de l'ingénierie informatique. Ici, le but est de montrer comment la multimodalité est qualifiée dans cette discipline ; d'abord selon un point de vue général, puis plus spécifiquement dans le contexte de la conception des énoncés multimodaux, en sortie du système.

2.4.1 Caractérisation des modalités de sortie dans les IHM

Les discussions dans ce domaine s'appuient sur des définitions diverses des notions de *'mode'*, de *'média'* et de *'modalité'* (*e.g.* Horchani, 2007; Jokinen & Raike, 2003; Le Bodic, 2005; Nigay & Coutaz, 1993). Les définitions proposées par les auteurs permettent généralement d'exprimer des propriétés spécifiques à l'approche formelle retenue par chacun. Les notions principales permettent de qualifier les éléments en présence. Ce sont : (1) les *'dispositifs techniques'* utilisés, (2) les *'organes perceptifs'* visés chez les utilisateurs et (3) le *'type de données'* véhiculé (linguistique, graphique, analogique). On peut noter également qu'ici le point de vue est centré sur le système et que l'objectif est de qualifier les moyens dont il dispose pour assurer la *'coopération entre les modalités'*[1]. On distingue deux problématiques :

- La *'fusion multimodale'* en entrée du système. Il s'agit d'interpréter les signaux en provenance de l'utilisateur et acquis par des dispositifs techniques différents. C'est un problème de *'reconnaissance'* qui a été étudié dès les premières études sur la multimodalité (Bolt, 1980) ;
- La *'fission multimodale'* en sortie du système. Il s'agit d'envoyer des signaux à l'utilisateur en utilisant plusieurs dispositifs techniques et de façon coordonnée. C'est une problématique de *'génération'* qui a émergé plus tardivement et qui se développe actuellement comme en atteste la *'feuille de route'* (« *roadmap »)* du séminaire international de Dagstuhl (Bunt, Kipp, Maybury, & Wahlster, 2003).

Ici, l'intérêt se porte plus spécialement sur la *fission*. Mais il est entendu que la notion de *couplage fonctionnel* interdit de considérer que ces problématiques seraient indépendantes.

A Le formalisme CASE et les propriétés CARE

Le formalisme CASE (Nigay & Coutaz, 1993) et les propriétés CARE (Coutaz & Nigay, 1994) permettent de définir l'espace problème de la conception des systèmes multimodaux et de qualifier les modalités utilisées.

[1] Dans cette expression, le terme « coopération » n'est pas entendu dans un sens psychologique. On peut l'entendre comme une notion large de « juxtaposition ». Différents niveaux de finesse sont possibles dans l'analyse des tâches qui définit le niveau de coopération. De même, le terme « modalité » doit être entendu dans un sens large, comme un moyen d'acquisition ou d'expression d'informations.

Le formalisme CASE est présenté dans le Tableau 2-7. Il a été conçu pour qualifier tout autant les entrées que les sorties du système. Il est basé aussi bien sur les données présentées que sur les propriétés physiques liées au dispositif technique. Les auteurs précisent que chacune des combinaisons (concurrente, alternée, synergique ou exclusive) peut être envisagée à différents niveaux d'abstraction, de la forme physique à la représentation sémantique. Le niveau d'abstraction est une troisième dimension qui s'ajoute aux deux dimensions croisées du Tableau 2-7 (*'mode de fusion'* et *'utilisation des modalités'*). Les quatre combinaisons résultantes correspondent, dans une perspective de développement, à quatre classes de systèmes multimodaux.

Tableau 2-7 : Représentation de l'espace de conception de la multimodalité (CASE)

		Utilisation des modalités	
		Séquentielle	**Parallèle**
Fusion / Fission	**Combinée**	ALTERNE	SYNERGIQUE
	Indépendante	EXCLUSIF	CONCURRENT

D'après Coutaz et Nigay (1994), les propriétés CARE permettent de décrire la communication entre l'humain et le système interactif. Elles reposent sur une définition de la *'modalité'* (notée : « M ») qui implique un dispositif physique (noté : « D ») et un langage d'interaction (noté : « L ») définit par un vocabulaire et une grammaire. Par exemple, la modalité « langage naturel » sera notée : $m_1(d_1, l_1)$ avec 'd_1' correspondant au haut-parleur et 'l_1' correspondant à un sous-ensemble d'une langue particulière pour laquelle est conçu le système.

Quand plusieurs modalités sont ainsi définies pour un système, il est possible de les combiner pour une tâche de présentation donnée de façon :

- ***Complémentaire*** : Différents types de contenus sont présentés sur des modalités distinctes ;
- ***Assignée*** : Un type de contenu est présenté sur une même modalité spécifique ;
- ***Redondante*** : Un type de contenu est présenté sur plusieurs modalités à la fois ;
- ***Equivalente*** : Un type de contenu est présenté indifféremment sur une modalité ou sur une autre (parmi plusieurs, mais pas nécessairement toutes).

B Composition des modalités

Cette caractérisation des modalités a donné lieu par la suite à des propositions pour la composition des modalités (Vernier & Nigay, 2000) dont l'objectif était de dépasser la notion de *'multimédia'*, vu comme une juxtaposition de supports, pour permettre la définition d'interfaces *'multimodales'* en sortie exploitant de façon dynamique la notion de *'média'* telle que définit ci-dessus. Cinq types de composition sont envisagés (issus de : Allen, 1983) et sont décrits selon cinq aspects différents (temporel, spatial, articulatoire, syntaxique et sémantique). La description résultante est présentée dans le Tableau 2-8.

Tableau 2-8 : Schémas de composition des modalités

Composition					
Temporelle	Anachronique	Séquentielle	Concomitante	Coïncidente	Parallèle / Simultanée
Spatiale	Disjointe	Adjacente	Intersectée	Imbriquée	Recouvrante
Articulatoire	Indépendante	Fissionnée	Fissionnée et dupliquée	Partiellement dupliquée	Dupliquée
Syntaxique	Différence	Complétion	Divergence	Extension	Jumelage
Sémantique	Concurrente	Complémentaire	Complémentaire et redondante	Partiellement redondante	Totalement redondante

A la suite de cette caractérisation des modes de composition des modalités, les auteurs cherchent à définir les critères de choix entre ces différentes possibilités. Ils précisent que ces critères sont indépendants du fait que ce choix revienne au système ou à l'utilisateur, dans l'interaction ; ou aux concepteurs, au cours du développement du système. Les pistes proposées pour identifier ces critères de choix sont issus des travaux de Bernsen (1994) sur des aspects descriptifs ou sur des critères d'utilisabilité qui peuvent être identifiés dans la littérature. La qualification des différents types d'effets que pourraient produire ces combinaisons sur le processus interactif (*i.e.* rôle de la composition des modalités dans l'interaction) a été envisagée comme une perspective nécessaire par les auteurs. Vernier et Nigay (2000) notaient que ce type de connaissance n'était pas présent dans la littérature psycho-ergonomique.

C Expressivité multimodale

Outre ces caractérisations des modalités, les bases de la problématiques des sorties multimodales en contexte interactif ont été abordées par Whittaker et Walker (1991).

Whittaker et Walker (1991) indiquent que le développement de systèmes interactifs multimodaux suppose de comprendre comment les différents média interagissent et qu'une théorie intégrant ces connaissances est encore manquante. Précisons que la même remarque a été faite plus récemment par l'un de ces auteurs et par d'autres (*e.g.* Oviatt, Coulston & Lunsford, 2004; Whittaker, 2003b). Il semble qu'elle reste d'actualité. Whittaker et Walker n'apportent pas de réponse à ce sujet, mais ils expriment les difficultés inhérentes à l'utilisation conjointe de différents médias.

Ils rappellent d'abord qu'en *communication médiatisée entre humains* l'hypothèse d'une *maximisation de la bande passante* (nommée *« bandwidth hypothesis »* dans Whittaker, 2003a) ne fonctionne pas. Autrement dit, il ne suffit pas d'ajouter un canal visuel pour améliorer la communication orale. Cela a été démontré par Chapanis et ses collaborateurs (Chapanis, Ochsman, Parrish, & Weeks, 1972; Chapanis & Overbey, 1974; Chapanis, Parrish, Ochsman, & Weeks, 1977). Les études de ces auteurs ne seront pas détaillées ici. Elles montrent que des *'tâches de communication distante'* entre humains sont correctement

réalisées à l'oral (par exemple, par téléphone) et que, dans cette situation, l'ajout de la modalité visuelle est neutre.

Whittaker et Walker (1991) remarquent quatre points :

(1) Des théories existent pour le traitement du langage naturel (analyse et génération), mais l'interaction entre textes et graphiques, en 1991, est faiblement théorisée. Ces points ont été abordés depuis, notamment dans le cadre des théories de l'apprentissage, qui sont évoquées en fin de chapitre 3 ;

(2) Les entrées et les sorties sont différentes. Une *théorie des sorties* est d'autant plus nécessaire que le système doit être capable de faire des choix sur ce qu'il présente ;

(3) De nouveaux médias (ordinateurs, téléphones, etc.) émergent et étendent les capacités des systèmes. Le besoin d'une approche générique se fait sentir. Cette remarque est d'autant plus vraie aujourd'hui ;

(4) Les recherches ont porté sur l'*interaction homme-machine'*, alors que l'ordinateur est très utilisé pour la *'communication homme-homme médiatisée'*. Depuis, cette remarque a été largement prise en compte. La réciproque peut parfois être formulée.

Les auteurs rappellent les principes des théories de la communication (*« grounding », Cf.* chapitre 1) et les propriétés des interfaces graphiques et des interfaces en langage naturel :

- ***Interfaces graphiques*** : Représentation continue des feedbacks, action physique directe, opérations visibles et réversibles ;

- ***Interfaces basées sur le langage*** : L'avantage général du langage est de permettre l'expression de *'relations abstraites'*, ce que les auteurs déclinent en un ensemble de caractéristiques : (1) référence par une *'description définie'* sans accès direct à l'objet, (2) références au *'contexte du discours'*, telles que les ellipses, (3) '*spécification temporelle'*, (4) *'quantification'*, (5) *'coordination'* des éléments du discours, (6) *'négation'*, (7) *'comparaisons'*, (8) *'modification des actions'*, (9) *'tri abstrait'*, (10) expressions *'conditionnelles'*, (11) *'causalité'*.

  On peut ajouter une remarque supplémentaire que ne font pas Whittaker et Walker. Le langage, en mode vocal, fait appel à l'audition, qui peut être considérée comme *« la sentinelle des sens »* (Hapeshi & Jones, 1992). Ainsi, il a également une fonction d'alerte qui peut avoir beaucoup d'importance dans l'interaction.

Whittaker et Walker (1991) résument que la parole s'avère utile pour l'interaction en temps réel, mais qu'elle a le défaut d'être éphémère (on parle d'*évanescence de l'oral*). Au contraire, l'affichage graphique permet de se passer de la structuration en tours de parole et d'organiser l'activité de l'utilisateur sur un mode parallèle. Lors d'un affichage graphique, une réponse immédiate n'est pas absolument nécessaire dans la mesure où les contenus présentés sont persistants, ce qui permet également de les structurer par des regroupements dans différentes zones. Pour ces raisons, la supériorité de la parole, mise en évidence par

Chapanis et ses collaborateurs, relève d'une vision trop simpliste. L'affichage graphique permet de mettre en commun les contenus.

En conclusion, ces remarques sont importantes, mais elles permettent seulement de mettre en évidence les *qualités* et *défauts* liés aux modes de présentation. Les théories doivent être adaptées pour permettre l'implémentation de systèmes capables de tirer partie de ces caractéristiques en fonction des besoins liés aux différentes situations, de combiner utilement différents médias et de mixer les modes de présentation selon les avantages de chacun et en vue d'une interaction conviviale.

2.4.2 Interaction multimodale dans les systèmes dialogiques

Des efforts récents ont été fournis pour faire *'la charnière'* (*« the hinge »*, voir Sun et al., 2008) entre entrées et sorties du système. Ces travaux ont un recouvrement important avec la problématique de l'utilisation des *'avatars'* dans la communication car ces *'personnages virtuels animés'* sont souvent vus comme une solution intéressante (*e.g.* Hernault, Piwek, Prendinger, & Ishizuka, 2008; Sun et al., 2008; Van Deemter et al., 2008) ; cela malgré des remarques de Whittaker (2003b) qui indique que dans des activités finalisées il est préférable de focaliser l'attention de l'utilisateur sur la *tâche finale*, alors que les avatars (ou la visiophonie, en CMC) focalisent l'attention sur la *tâche de communication*. Ainsi, ces travaux étendent les possibilités multimodales des systèmes. Ils exploitent le plus souvent des algorithmes disponibles dans les systèmes préexistant et les combinent pour proposer des solutions innovantes. La logique prédominante est celle de l'analyse grammaticale (*« parsing »*) dont la pertinence n'est pas toujours questionnée.

A L'interface « Speech-in List-out » (SILO)

Une interface générique de DHM qui exploite la présentation visuelle de certaines informations a été proposée par Divi et al. (2004). Cette interface est nommé *'Parole-en-Entrée Liste-en-Sortie'* (*« Speech-In List-Out »*, ou 'SILO'). Les auteurs proposent une expérimentation qui montre une bonne robustesse de la reconnaissance vocale dans leur système et qui indique que la *'charge cognitive'* (*« cognitive load »*) de l'utilisateur est réduite en comparaison à une interface basée sur des menus.

Le principe de fonctionnement du système correspond à une logique de remplissage de formulaire (*'Système à gabarits', Cf.* Tableau 2-1). Il est basé sur une reconnaissance vocale qui n'implique pas une conversion textuelle (*« speech-to-text »*) syntaxiquement correcte de la demande de l'utilisateur. Les mots reconnus sont placés dans un *'sac de mots'* (*« bag of words »*) et organisés selon leur fréquence d'apparition dans les requêtes des utilisateurs. Cela permet de lancer une recherche (requête à la base de données) pour identifier un ensemble de réponses candidates. Ces réponses sont ensuite présentées textuellement à l'utilisateur, sous la forme d'une liste. Il est alors plus facile pour l'utilisateur de choisir la solution qui l'intéresse. Ainsi, ce fonctionnement est une extension de la logique du *'n-best'*,

qui permet également de produire des listes de candidats accompagnés chacun d'un degré de confiance. Mais le *'n-best'* est calculé directement au niveau du modèle de reconnaissance vocale, alors que dans ce cas il s'agit d'une logique plus large qui implique l'ensemble des composants de l'application (reconnaissance vocale, base de donnée et gestionnaire du dialogue).

A partir de ces principes de fonctionnement, les auteurs ont réalisé des expériences pour évaluer l'efficacité du système. Ils ont d'abord travaillé sur la pertinence des requêtes qui était évaluée par une équipe de juges indépendants. Cette première évaluation a permis de montrer que l'amélioration obtenue par rapport à une reconnaissance vocale classique correspondait à l'amélioration quand le bruit de fond est plus bas de 5 dB. Les auteurs ont alors réalisé une seconde évaluation plus générale de l'interface SILO, en termes de performance. Dans une première expérience, ils ont comparé l'interface SILO à une interface textuelle standard dans un outil de recherche de chanson (*'MediaFinder'*). L'interface textuelle était légèrement plus efficace de 5% à 0,5% en fonction de la taille de l'index (de 6828 à 54 chansons). L'interface textuelle ayant un score de 99% environ – quelle que soit la taille de l'index – l'interface SILO s'avérait d'une bonne efficacité. Dans une seconde expérience, les auteurs ont renouvelé la même comparaison dans un environnement de conduite automobile. Les résultats ont montré que quand l'interface SILO était utilisée le nombre d'erreurs de navigation dans la conduite était inférieur de 20,7% par rapport à l'interface textuelle et le temps nécessaire pour trouver une chanson était inférieur de 28,6%. Cependant, les auteurs notent que la performance de conduite était plus basse quelle que soit l'interface. Ils en ont conclut que l'interface SILO imposait une *'charge cognitive'* (*« cognitive load »*) significativement plus basse en comparaison à l'interface textuelle.

B Choix dynamique de la modalité

Quelques travaux récents portent sur le développement d'algorithmes de génération multimodale destinés à permettre le choix dynamique de la modalité ou de la combinaison de modalités à utiliser (*e.g.* Bachvarova, Van Dijk & Nijholt, 2007; Guhe, 2007; Keizer & Bunt, 2007b). Ces travaux s'appuient sur les principes de modélisation cognitive disponibles. Comme le note Guhe (2007) : *« Les principaux problèmes sont que les théories unifiées existantes de la cognition ne détaillent pas l'emploi du langage. »* (p. 61, traduction libre). Pour cette raison, cet auteur se reporte sur le modèle de Levelt (1989). Il prend en compte les différents niveaux de processus proposés par Levelt pour la génération des structures sémantiques : (1) construction, (2) sélection, (3) linéarisation et (4) génération. Guhe (2007), discute les relations entre les représentations, d'ordre verbal, situées à ces différents niveaux, et d'autres représentations, notamment spatiales. A partir de ces éléments, il propose une réflexion sur la localisation de la fission multimodale dans la chaîne de traitement. Ses arguments indiquent que le niveau de précision impliqué par la micro-planification, en bout de chaîne, entraîne des relations modales qui obligent à ne faire la fission que très tard dans la série de processus. Il indique dans sa conclusion que la conceptualisation sémantique n'est

certainement pas suffisante pour déterminer le comportement multimodal. En effet, des relations, par exemple d'ordre spatial, peuvent également avoir du poids. Selon ce point de vue, la modélisation d'une chaîne modulaire de traitement entraîne nécessairement une complexité importante.

- ***Ontologie des modalités***

Bachvarova, Van Dijk et Nijholt (2007) proposent, dans la lignée des travaux sur la caractérisation des modalités, de décrire une *'ontologie des modalités'* (*« modality ontology »*) qui tente de réconcilier divers aspects, pour dresser un tableau pour des recherches futures. L'ontologie est basée sur une différence entre le *'contenu'* d'une modalité, qui correspond à l'information à présenter, et sont *'profil'*, qui correspond à sa nature. Le profil est divisé en trois niveaux (Tableau 2-9).

Tableau 2-9 : Les trois niveaux de description du *'profil'* des informations

(1) Le *'niveau de la présentation de l'information'*
indique quel type de contenu peut être associé à chaque modalité. Il suppose une opération d'appariement (*« mapping »*). Ce contenu est *linguistique* ou *analogique*. Les auteurs associent des propriétés à chacun de ces deux types de représentation. Ils indiquent que les représentations linguistiques (texte ou parole) ne précisent pas les spécificités des objets et qu'elles sont abstraites et focalisées sur le contenu. Au contraire, les représentations analogiques véhiculeraient des informations visuelles ou spatiales qui ne sont ni focalisées sur le contenu, ni abstraites.
(2) Le *'niveau de la perception'*
indique quels processus perceptifs et cognitifs peuvent être réalisés à partir de chaque modalité. La modalité est *auditive*, *visuelle* ou *tactile* (*« haptic »*). La dimension temporelle est vue comme importante à ce niveau et donne lieu à une distinction entre modalité *statique* et *dynamique*. Pour combiner les modalités, les auteurs renvoient au *'principe de modalité'* tel qu'il est établit par Moreno et Mayer (1999). Sur cette base, les auteurs renvoient à l'évitement des *surcharges cognitives* permis, d'une part, par l'attribution de la modalité visuelle aux informations analogiques (*e.g.* animations) et, d'autre part, par l'attribution de la modalité auditive aux informations linguistiques (*i.e.* parole).
(3) Le *'niveau des dépendances structurelles'*
indique les règles qui déterminent la syntaxe d'association des modalités entre elles. Les auteurs renvoient à Bernsen (1994) qui indique que des modalités peuvent être *dépendantes* ou *indépendantes*.

L'idée centrale de cette approche est que le sens est déterminé par ces trois niveaux. Les auteurs indiquent que ces niveaux permettent de représenter les aspects spécifiques qui ne sont pas déterminés par la représentation sémantique.

Au sujet des contenus linguistiques, ils indiquent que le langage écrit est indépendant de la situation alors que le langage parlé est supposé servir la communication située. Cette précision permet de compléter le point de vue apporté par les travaux de Moreno et Mayer (1999), utilisés comme référence en psychologie, mais qui ne précisaient pas ces aspects. Ces auteurs souhaitent développer une approche comportementale complète. Ils consacrent la suite de l'article aux principes de fonctionnement technique qui permettent de lier le *'contenu'* au *'profil'* dans un système, aspect non développé ici.

C Formalisation des fonctions dialogiques

Par ailleurs, une approche différente est proposée par Keizer et Bunt (2006; 2007a; 2007b). Cette approche exploite la taxonomie de Bunt (1994, *Cf.* Figure 2-4). Elle consiste à baser les choix d'action et/ou d'interprétation sur les diverses fonctions des actes qui composent les énoncés. Les auteurs précisent que la prise en compte de l'aspect multifonctionnel de chacun des actes n'est pas simple et qu'elle peut impliquer des redondances entre les actes, ou au contraire des conflits, que l'algorithme doit être en mesure de prendre en compte. La solution proposée exploite l'approche multi-agent (voir Wooldridge, 2002) qui permet de gérer cette complexité en fonction d'éléments contextuels (contexte cognitif, sémantique, social et linguistique). Les agents qui composent le système sont chargés de la gestion des contraintes liées à ces différents éléments du contexte.

Mais si le système proposé est en mesure d'évoquer des éléments multimodaux (par exemple, des boutons présentés à l'écran), les auteurs ne précisent pas quelle taxonomie est utilisée pour classifier les modalités. Par ailleurs, le *'contexte cognitif'* décrit par les auteurs inclut l'état de compréhension de l'utilisateur au regard de l'avancement de la tâche (*e.g.* compréhension ou incompréhension d'un feedback). La tâche est modélisée sous la forme d'actes de dialogue, certes finement interdépendants, mais qui n'intègrent pas, par exemple, la dimension perceptive qui a été évoquée dans l'article de Guhe (2007). Ainsi, il semble toujours nécessaire pour ces approches, de converger. En outre, le *'contexte social'* modélise plus particulièrement les marques de politesse dans la conversation et n'intègre pas des aspects relationnels plus complexes. Pour ces différentes raisons, il semble que la modélisation de l'interlocuteur puisse encore être améliorée par un élargissement du *'champ de conscience'* implémenté dans le système.

2.4.3 Conclusion

A travers ces différents travaux, on constate que la description des modalités a progressé depuis le développement des longues listes taxonomiques de Bernsen (1994). Cela permet d'implémenter des systèmes capables de faire le *« mapping »* (*'appariement'*) nécessaire pour présenter des informations sur plusieurs dispositifs techniques de façon parallèle et synchronisée. On doit cependant remarquer qu'il y aurait une certaine rigidité dans une opération d'appariement qui ne prendrait en compte que des caractéristiques descriptives. Notamment, l'association entre matériel linguistique et modalité auditive qui a été suggérée par Bachvarova, Van Dijk et Nijholt (2007) sur la base des résultats de Moreno et Mayer (1999) est insuffisante dans le contexte d'un système de DHM où tout le matériel à présenter est linguistique. La prise en compte stricte de ce résultat indiquerait que les énoncés verbaux de ces systèmes sont suffisants et qu'aucun effort de conception n'est nécessaire.

Les travaux en ingénierie des sorties multimodales ont permis d'exploiter le potentiel offert par les analyses faites en sciences-humaines (*e.g.* Divi et al., 2004 ; Keizer & Bunt, 2007a). Ces travaux, ainsi que ceux sur la caractérisation des modalités, ouvrent des possibilités pour

le développement des interfaces de communication en langage naturel. Mais ces auteurs remarquent fréquemment que les effets des modes de communication sont mal connus et que les théories psychologiques ne répondent pas à tous les questionnements qui se posent dans le cadre de la conception d'un système.

Comme le notent Vernier et Nigay (2000), les études manquent sur la complémentarité des modalités. C'est notamment le cas dans le contexte du DHM où il est nécessaire d'identifier les différents types d'effets qui peuvent survenir. Sur ce point, Jokinen et Raike (2003) indiquent : *« Les systèmes d'interaction naturelle ne doivent pas requérir des utilisateurs qu'ils apprennent des tâches physiques compliquées (pointage et cliquage) ni qu'ils réalisent des traitements cognitifs extensifs (systèmes de menus), mais doivent leur fournir des feedbacks clairs et immédiats sur leurs actions, suivant des patterns comportementaux que des humains emploieraient dans des situations similaires »* (p. 9, traduction libre). La difficulté provient sans doute du fait qu'il est nécessaire de croiser une analyse fonctionnelle intégrée, telle que celle de la pragmatique, et une analyse descriptive/structurelle qui permette à la fois de qualifier les *'modes de présentation'* liés aux dispositifs techniques et les *'modalités perceptives'* humaines responsables du captage des messages.

Par ailleurs, ces auteurs en sciences de l'ingénieur notent les limites émanant de l'isolement des modèles psychologiques. On trouve en effet des modèles linguistiques (*e.g.* Levelt, 1989), des modèles de l'apprentissage (*e.g.* Mayer, 2005) et des modèles intégrés de la cognition (ou *'de l'esprit'*. *Cf.* Anderson et al., 2004), mais ces modèles manquent de cohérence dans leur principes (cohérence *'inter-modèles'*) et ne sont pas porteurs d'un discours homogène en direction des communautés extérieures à celle de la psychologie. La partie suivante va permettre de présenter le point de vue développé par la psychologie spécifiquement sur les principes de présentation multimédia des informations.

2.5 Synthèse

Finalement, le développement des services vocaux est une réalité désormais quotidienne (ex : répondeur) et les pistes d'amélioration sont nombreuses. L'évolution des technologies vocales a permis d'accroître la taille du vocabulaire reconnu, d'améliorer les méthodes de traitement sémantique et linguistique et d'introduire des méthodes d'interaction plus riches (Juang & Rabiner, 2004). Ces techniques permettent de développer des systèmes de plus en plus flexibles, utiles dans le cadre d'une utilisation commerciale. Mais les auteurs notent cependant que la conception de systèmes adoptant un comportement intelligent reste un challenge important pour l'avenir (Juang & Rabiner, 2004; Mc Tear, 2002).

Minker et Néel (2002) notent : *« on commence à prendre en considération la complémentarité de la parole avec d'autres modes de perception ou de production (gestuel ou tactile, notamment). Ceci impose une conception radicalement différente des systèmes afin d'en permettre une intégration efficace. »* De même, certains auteurs notent le manque de modèle

uniforme et complet pour la gestion de l'interaction multimodale (*e.g.* Oviatt, Coulston & Lunsford, 2004). Il est donc nécessaire d'étudier ces aspects pour concevoir des systèmes au comportement plus naturel et plus efficace.

Chapitre 3 L'étude empirique des énoncés en contexte interactif

L'objectif de ce chapitre est de présenter les travaux empiriques qui ont permis d'étudier les effets des énoncés du système dans le cadre de l'utilisation d'un système interactif.

Dans un premier temps, une partie consacrée au positionnement du problème (ci-dessous) va permettre de donner un aperçu du contexte de la recherche dans ce domaine. Les parties suivantes portent sur l'état de l'art en matière de *conception des énon*cés *des systèmes interactif*s, d'abord dans le contexte du DHM, puis dans le domaine de la psychologie des apprentissages.

3.1 Une thématique, des problématiques

Il y a une difficulté intrinsèque à présenter le thème de la *conception des énoncés du système*, qui tient au fait que cette thématique ne peut être rapprochée exclusivement, ni d'une discipline particulière, ni d'une problématique unique. Globalement, le thème se rattache aux *sciences de l'information* qui renvoient, comme l'indique Wilson (1997), à une vaste palette de disciplines. Parmi elles, les *sciences cognitives* couvrent notamment la *linguistique*, la *psychologie* et les *sciences de l'ingénieur*. Dans ce champ, différents points de vue peuvent être isolés, émanant par exemple de couplages disciplinaires (*'linguistique-ingénierie'*, *'psychologie-ingénierie'*, *'linguistique-philosophie-mathématiques'*, etc.). Chacun de ces points de vue connaît, au cours du temps, des évolutions spécifiques dans ses perspectives, liées aux progrès techniques et à la diversification des terrains d'application. On peut constater des déséquilibres liés à des difficultés méthodologiques spécifiques à chaque discipline, ainsi qu'aux relations interdisciplinaires, souvent agrémentées de divergences terminologiques qui sont parfois insurmontables. La synthèse n'est donc pas simple.

3.1.1 Approche multidisciplinaire

Tout d'abord, les mêmes sujets sont souvent traités dans des perspectives différentes par les *linguistes*, les *psychologues* et les *informaticiens*. Par exemple, la question de l'utilisation des pronoms et des *ellipses pronominales* peut être abordée par les linguistes sous l'angle de leurs emplois dans des conversations entre humains, analysées sur la base de corpus d'enregistrements de dialogue. Des psychologues s'intéresseront plutôt, suivant des méthodes plus expérimentales, aux processus cognitifs liés à la production et la

compréhension des ellipses, ou à l'acquisition de ces processus au cours du développement. Des informaticiens chercheront quant à eux à modéliser une *'grammaire universelle'* (hypothétique) permettant de faire produire des ellipses « naturelles » à un système automatique, quelle que soit la langue qu'il utilise.

Or, les modèles conceptuels sous-jacents diffèrent et peuvent souffrir d'une certaine étanchéité qui ne permet pas toujours la mise en commun des résultats et des enseignements. Notamment, dans l'optique de la *formalisation computationnelle du langage* (voir, par exemple, Jackendoff, 2007), les collaborations interdisciplinaires ont d'abord concerné *informaticiens* et *linguistes*, qui ont pu faire émerger un *'terrain commun'* conceptuel interdisciplinaire (en priorité sur des bases grammairiennes et structuralistes, et également sur les bases de la philosophie analytique, *e.g.* Searle & Vanderveken, 1986). Il s'ensuit qu'une reprise de cette problématique par des *psychologues* donne l'impression d'une redite pour les tenants des autres disciplines, qui peuvent ainsi passer à coté d'un éclairage nouveau, ou complémentaire.

De même, la question de la *multimodalité* échappe en grande partie à la linguistique et elle est conçue différemment en informatique et en psychologie ; ou encore, le problème des *apprentissages*, largement étudié par les psychologues, est totalement réinterprété par les informaticiens (sur des bases avant tout statistiques), etc.

Ainsi, pour chaque problème, des recoupements interdisciplinaires sont nécessaires afin d'embrasser une perspective qui ne soit pas trop limitée (voir Dahlbäck, 2003).

3.1.2 Questions transversales

Quelle que soit la discipline, des problématiques diverses peuvent être isolées. Pour en donner une idée, les oppositions conceptuelles les plus courantes ont été présentées dans le Tableau 3-1 sous forme de couples. Il n'y a pas nécessairement d'opposition dans ces couples et les différents problèmes qu'ils soulèvent sont souvent étudiés de front. Une diversité de points de vue peut être adoptée et tout questionnement doit y être situé.

Ici, l'attention se porte sur le *'DHM vocal'* et la perspective est celle d'une extension à des *'sorties' 'multimodales'*. On s'intéressera à leurs impacts en termes d'*apprentissage humain* (tant du point de vue de la *'représentation'* que de celui de l'*interaction'*) et à leurs '*répercussions sur'* la charge cognitive de l'utilisateur.

Tableau 3-1 : Quelques problématiques en lien avec la conception des énoncés

***'Dialogue Homme-Homme'* (DHH) *vs. 'Dialogue Homme-Machine'* (DHM)**
Le DHH a d'abord été utilisé comme base de connaissance pour la conception de systèmes de DHM ; et ces deux champs ont fait l'objet de comparaisons (*e.g.* Kennedy, Wilkes, Elder, & Murray, 1988). Par ailleurs, chacun de ces deux champs est un domaine entier. Notamment, le DHH est dépendant des outils utilisés pour médiatiser la communication (domaine des CMC, pour *« Computer Mediated Communication »*) et il a fait l'objet d'un grand nombre d'études (pour une revue, voir Whittaker, 2003a). Le DHM pose, quant à lui, certains problèmes spécifiques auxquels ce chapitre doit sensibiliser ;
***'Entrées' vs. 'Sorties'* (*« inputs » vs. « outputs »*)**
Les sorties du système (énoncés produits vers l'utilisateur) renvoient à la chaîne de traitement qui permet de les produire. Il s'agit d'une problématique à part entière pour l'informatique. Ces énoncés produisent des effets chez l'utilisateur et sur les réponses qu'il adressera en retour vers les entrées du système, qui constituent également une problématique à part entière (*e.g.* reconnaissance vocale). Ces questions ne peuvent être totalement isolées puisqu'elles s'influencent mutuellement. On peut ajouter que la question des *'modes de communication'* (*e.g.* Le Bigot, Rouet & Jamet, 2007; Le Bigot et al., 2007) considère des *patterns entrées-sorties* fixés (*e.g. 'clavier-écran'* ou *'micro-haut-parleur'*) et étudiés comme un tout en psycho-ergonomie. Par ailleurs, la caractérisation des langages d'interaction est également traitée en ingénierie des interactions homme-machine en entrée (*e.g.* Nigay et Coutaz, 1993) et en sortie (*e.g.* Vernier, 2000) ;
'Vocal' vs. 'Multimodal'
Les systèmes vocaux et graphiques ont d'abord été conçus isolément avant qu'une convergence technique ne soit possible, donnant lieu aux systèmes multimodaux. Au-delà des questions techniques complexes qui accompagnent cette convergence, les principes d'interaction avec l'utilisateur et d'apprentissage des contenus ne se posent pas dans les mêmes termes. Les questions techniques sont spécifiques. L'analyse des interactions doit être globale et relève d'une analyse fonctionnelle. De plus, la multimodalité génère une diversité de problématiques dont seules les plus pertinentes pour la thèse sont évoquées, plus loin, dans ce document ;
'Représentation' vs. 'Interaction'
Dans le domaine strict des interfaces vocales, la question des représentations des utilisateurs sur le système a d'abord été étudiée pour isoler leur impact sur les verbalisations des utilisateurs (*e.g.* Kennedy, Wilkes, Elder, & Murray, 1988 ; Amalberti, Carbonnell, & Falzon, 1993). Cette question est liée au problème des *entrées*, question cruciale étant donné les taux d'erreurs de reconnaissance vocale inhérents à ces systèmes (Zoltan-Ford, 1991). La question de la modélisation des interactions joue sur un plan différent lié à la modélisation des états du système (*Cf.* section précédente). Les études peuvent être focalisées sur l'un ou l'autre de ces aspects, ou les englober ;
***'Influence sur' vs. 'Indicateurs de'* charge cognitive**
La question de la charge cognitive de l'utilisateur est récurrente dans la littérature. Les études peuvent être classées dans des catégories distinctes selon qu'elles cherchent à identifier les facteurs extérieurs (sorties du système) qui influencent la charge (*e.g.* Baber et al., 1996; Oviatt, Coulston & Lunsford, 2004) ou, au contraire, les indicateurs en provenance de l'utilisateur (entrées du système) qui permettraient de diagnostiquer qu'il est « en surcharge » à un temps « t » (*e.g.* Berthold & Jameson, 1999; Jameson et al., 2006; Yin & Fang, 2007) ;
'Apprentissage humain' vs. 'Apprentissage machine'
Les systèmes automatiques sont aujourd'hui capables d'apprentissage et ce domaine est en pleine expansion (*e.g.* Baudoin, 2007). Ces apprentissages peuvent porter sur chacun des domaines évoqués précédemment. De même, l'humain en interaction avec le système est, dans tous les cas, en situation d'apprentissage, tant du point de vue des interactions que de la compréhension/mémorisation du contenu. Mais ces domaines ne se recoupent que rarement.

3.2 L'étude des énoncés en DHM

En DHM vocal, la difficulté liée à la conception des énoncés du système, en sortie, tient à l'impossibilité de concevoir une reconnaissance vocale parfaite (*e.g.* Deng & Huang, 2004; Lippmann, 1997), en entrée. La nature variable de la parole, et la nécessité de définir un vocabulaire limité (quelle que soit sa taille), font qu'il est impossible d'atteindre la perfection car le recouvrement de ces deux ensembles (parole productible *vs.* parole compréhensible) n'est jamais total. De ce fait, il est nécessaire de modeler (*i.e.* influencer) le comportement de l'utilisateur par différentes techniques.

Dans un premier temps, on a constaté la variabilité des comportements verbaux des utilisateurs et les études se sont focalisées sur la mauvaise représentation qu'ils avaient des capacités du système. L'hypothèse était qu'une représentation cohérente avec la réalité des capacités du système permettrait une meilleure adaptation. Les résultats (*e.g.* Amalberti, Carbonnell, & Falzon, 1993 ; Kennedy, Wilkes, Elder, & Murray, 1988) montrent en effet que, pour un comportement identique du système, les commandes de l'utilisateur sont plus simples à reconnaître (car plus courtes et avec un vocabulaire moins riche) quand il pense parler à une machine que quand il pense parler à un humain. D'après Amalberti et al. (1993), quand ils parlent à une machine, les participants utiliseraient le système comme un outil. Des études plus récentes (Amiel, 2005 ; Le Bigot, 2004) montrent en effet que les différences portent sur les indices de collaboration mis en avant par Clark et ses collaborateurs. Les utilisateurs ont une attitude moins collaborative quand ils pensent parler à une machine. Cette question est toujours en discussion aujourd'hui dans les termes d'une humanisation des systèmes de DHM (*« human-like SDS »*, Edlund, Gustafson, Heldner, & Hjalmarsson, 2008).

3.2.1 Principes de gestion de l'interaction pour le DHM

A Influencer l'utilisateur

Le fait d'influencer l'utilisateur (*« shaping user »*, *Cf.* Leiser, 1989; Ringle & Halstead-Nussloch, 1989; Zoltan-Ford, 1991) consiste à l'origine à lui permettre de réutiliser le vocabulaire et les structures syntaxiques utilisés dans les énoncés du système (en sortie) lorsqu'il produit une commande (en entrée). Cette technique permet de placer les contraintes sur l'utilisateur, dans la mesure où celui-ci a de meilleures capacités d'adaptation.

Cette idée est inspirée de *'l'effet de durée de la parole'* (Matarazzo, Weitman, Saslow, & Wiens, 1963) qui montre que, pendant une interview, la longueur des réponses est corrélée positivement avec la longueur des questions. Ce type d'effet existe également dans la longueur des pauses, l'intensité de la parole et autres (Leiser, 1989). Ainsi, la solution proposée pour les systèmes de DHM consiste à influencer l'utilisateur subrepticement (*« covertly »*) plutôt qu'explicitement (*« overtly »*). Les énoncés du système sont conçus en

cohérence avec le vocabulaire reconnaissable, ce qui permet de bien guider l'utilisateur sans trop alourdir le dialogue. Cet effet a été validé par Zoltan-Ford (1991) dans le cadre d'un système de DHM. Les participants s'alignaient sur la longueur des énoncés du système indépendamment de leur expertise préalable en informatique. Les énoncés courts avaient une influence plus grande que les énoncés conversationnels, plus longs. Mais comme le remarque cet auteur, bien que cette technique améliore l'utilisabilité du système la variabilité du comportement des utilisateurs existe toujours. Même dans la meilleure condition (énoncés courts) l'influence du système était totale avec seulement 51,4 % des participants. Les autres utilisateurs introduisaient des termes non reconnus par le système. Ce résultat est important et, depuis, cette technique est appliquée dans la conception des dialogues des systèmes conçus pour le téléphone.

Par ailleurs, l'influence du comportement du système ne porte pas seulement sur l'aspect linguistique. Par exemple, Johnstone, Berry, Nguyen et Asper (1994) ont comparés l'organisation des tours de parole en DHH et en DHM. Pour la condition DHM, ils ont développé un protocole incluant deux magiciens d'Oz (*Cf.* chapitre 5). Le premier magicien écoutait le participant et sélectionnait une réponse prédéfinie qui était communiquée au second magicien, lequel choisissait le feedback à envoyer au participant. Ce protocole était conçu pour reproduire au mieux le comportement d'automate du système, mais il allongeait également un peu les temps de réponse. Cette particularité a eu pour conséquence de réorganiser le comportement des participants de façon importante puisque les pauses provoquaient des hésitations et des interprétations particulières. Le résultat est que pour la même tâche (saisie de numéros) 15 tours de parole étaient nécessaires en DHH et seulement 7 en DHM, le nombre de mots variant en sens inverse. Les auteurs expliquent dans la discussion que c'est le *processus de grounding* dans son ensemble qui était modifié, mais que la compréhension (le terrain commun) était construite aussi bien dans un cas que dans l'autre. Ils estiment, contrairement à Amalberti et al. (1993) que le DHM n'est pas fondamentalement différent du DHH et qu'il ne s'agit probablement que d'un problème de richesse comportementale. Ils soulignent par là l'importance du type d'interaction mis en place et relèguent au second plan le rôle des connaissances sur l'interlocuteur.

Dans le même sens, une tentative assez récente (Wilkie, Jack & Littlewood, 2005) proposait d'utiliser différentes stratégies de politesse pour reprendre un dialogue après une interruption involontaire. Mais ces auteurs en ont conclu que l'interruption produisait un effet négatif quelle que soit cette stratégie. La stratégie employée ne peut se reposer simplement sur l'ajout ou la suppression de différents contenus.

B Impact des "restrictions de syntaxe" dans les énoncés

L'expression *« restriction de syntaxe »* ne désigne pas, comme on peut le penser a priori, des réductions de la complexité des énoncés, auxquels on aurait retiré leur syntaxe. Au contraire, il s'agit d'indications supplémentaires destinées à informer l'utilisateur au sujet des capacités de compréhension limitées du système. Ce sont des prompts qui informent de la restriction

des capacités de compréhension du système. L'ajout de ces *'restrictions de syntaxe'* a été testé par Bubb-Lewis et Scerbo (2002) qui ont montré l'impact positif de l'ajout d'informations contextuelles. Celles-ci portaient sur les commandes disponibles et précisaient plus ou moins l'état en cours dans le dialogue. Dans cette expérience, quand le système fournissait moins d'information, les participants prenaient moins d'initiatives, de sorte que le contrôle du dialogue revenait au système. Dans le même sens, Murray, Jones et Frankish (1996, expérience 1) ont également ajouté des restrictions de syntaxe qui ont amélioré le processus collaboratif et le naturel de la communication. Mais dans leur expérience, les restrictions de syntaxe complètes étaient temporellement coûteuses et des indications partielles étaient préférables. On peut en conclure que dans les interfaces purement vocales les informations fournies à l'utilisateur sont utiles pour contrôler l'interaction, mais qu'elles sont également coûteuses du point de vue temporel.

Sheeder et Balogh (2003) ont travaillé sur l'impact de l'énoncé initial, en début de dialogue (*'prompt d'ouverture'* : *« initial prompt »*) dans un système développé pour l'industrie (technologie Nuance®). Ils montrent que l'ajout d'un exemple en début de dialogue, sur un mode naturel, permet d'améliorer significativement les routages corrects des appels (exemple : *« Bienvenue sur le service X. Vous pouvez me demander des choses telles que ''Combien de minutes ai-je utilisé ?'' ou ''J'aimerais un paiement automatique.'' Alors, comment puis-je vous aider au sujet de votre compte ? »* p. 105, traduction libre). L'amélioration du routage a eu lieu alors que ni la durée du dialogue, ni le nombre de mots prononcés n'a été impacté. Ces indications ont simplement donné aux participants les moyens de mieux communiquer leurs buts. On peut noter que, cette fois, ce sont les connaissances des participants qui ont été impactées ; connaissances de nature procédurale. Si une limite doit être notée dans cette étude, on peut préciser que la relation étudiée porte sur la réponse immédiate des utilisateurs à la suite de l'énoncé système. Il pourrait être intéressant de vérifier l'impact sur la suite du dialogue, dans un dialogue incluant plus d'étapes.

C Impact de la « visualisation » (métaphore du bureau)

Les énoncés du système peuvent aussi fournir des contenus d'une autre nature. Récemment, Howell, Love et Turner (2006) ont réalisé une étude qu'ils présentent dans un article dont le titre peut être trompeur (*''La visualisation améliore l'utilisabilité des services vocaux pour téléphone mobile''*). Ici, le terme *'visualisation'* ne désigne pas un affichage visuel des informations, mais l'utilisation de *métaphores visuelles* dans des énoncés vocaux pour décrire la structure hiérarchique des menus. L'interface était entièrement vocale. Ces *« métaphores visuelles »* consistent à décrire des objets matériels pour présenter les menus du service de façon à utiliser les capacités de représentation spatiale des participants. Dans cette expérience, deux versions étaient testées (trois en incluant la condition *'menu classique'*) :

(1) *Métaphore du bureau Windows* : L'une était issue de l'univers du PC. Les menus étaient nommés *'dossiers'*, *'fichiers'*, *'corbeille'*, etc.

(2) *Métaphore de l'environnement de travail* : L'autre était inspirée d'un environnement réel de travail. Les menus étaient assimilés à des pièces dans lesquelles l'utilisateur se déplace.

Cette étude montre que l'utilisation de métaphores peut améliorer l'utilisabilité du système. La métaphore du bureau Windows était la plus appréciée du point de vue des évaluations subjectives, excepté la sensation de vitesse. En effet, les participants pouvaient réutiliser des connaissances acquises dans l'univers du PC, qui les aidaient à comprendre les choix qui leur étaient proposés. Les auteurs supposent que la sensation de lenteur provient de la comparaison avec l'ordinateur classique où l'affichage graphique peut être exploré plus rapidement. La métaphore de l'environnement de travail était également évaluée plus positivement que les menus vocaux classiques et la sensation de vitesse était la plus élevée dans cette condition. En revanche, du point de vue de la performance, aucune amélioration n'a été obtenue dans ces deux conditions. En effet, les participants semblaient comprendre plus vite les choix qui se présentaient à eux, leur confiance augmentait, ils exploraient plus spontanément le service et ils interrompaient plus fréquemment le système. Mais du fait de ce comportement, ils généraient des erreurs dont le rattrapage était coûteux (durée du dialogue et nombre de tours de parole). Globalement, ce résultat n'est pas négatif, au contraire, puisqu'il guidait correctement les participants.

Dans leurs conclusions, les auteurs notent la tendance de nombreux participants à construire une représentation visuelle des menus quelle que soit la condition. Cette remarque est importante car elle montre que les énoncés des systèmes ne sont pas traités sur un mode uniquement verbal. Les schémas que construisent les participants sont plus complexes.

D Utilisation de la modalité visuelle dans le DHM

L'affichage visuel des contenus fournit également une aide intéressante. Murray, Jones et Frankish (1996), dans l'article déjà évoqué, présentent une seconde expérience dans laquelle ils utilisent l'affichage visuel dans le cadre du DHM. Dans cette expérience, les contenus présentés visuellement portaient sur les champs à compléter dans une tâche de remplissage de formulaire. Deux types d'indications visuelles étaient testés. Le premier indiquait le *'nom du champ'* à remplir ; par exemple : « jour de la semaine ». Le second listait *'toutes les options'* possibles ; par exemple : « lundi, mardi, etc. » Dans ces deux conditions, quand un mot était prononcé le feedback consistait à l'afficher dans le champ approprié. Les résultats ont montré l'intérêt de l'affichage visuel puisque 15% des erreurs n'étaient pas corrigées quand l'affichage visuel était présent (quelle que soit la condition) contre 25% dans la condition purement auditive. Entre les deux conditions visuelles, c'est le listage des options qui conduisait à la meilleure performance. En effet, quand les options étaient listées, l'erreur la plus fréquente consistait à substituer une option à une autre. Quand seul le nom du champ était indiqué, il arrivait plus fréquemment que les participants utilisent des items hors vocabulaire, qui n'étaient donc pas reconnus. La limite qui peut être notée dans cette expérience est que seules les indications de « restriction de syntaxe » sont présentées

visuellement. Pour aller plus loin, il est possible de chercher à catégoriser les contenus que le système doit présenter (voir chapitre 2). La manipulation d'ensemble des différentes catégories peut fournir plus d'informations.

Dans ce sens, une proposition de Yin et Zhai (2006) consiste à étendre l'interface de sortie des systèmes de DHM – basée uniquement sur la génération sonore de la voix – en ajoutant un affichage visuel. Ces auteurs ont réalisé trois expériences différentes qui montrent l'intérêt d'une telle extension. Dans la première, ils ont choisi un service simple pour lequel ils ont dupliqué l'ensemble des contenus en visuel et en vocal ; et ils ont testé ce service avec un groupe de six participants auxquels ils ont simplement demandé de commenter leur interaction avec le service. Les commentaires des participants étant positifs, ils ont alors conduit une seconde expérience, avec 16 participants, dans laquelle ils ont fait des mesures de performance. La durée moyenne des dialogues passait de 75 secondes à 50 secondes, environ (lecture graphique, valeurs non communiquées dans l'article). Le taux d'erreur (*'nombre de tours de dialogue avec erreur'* divisé par *'nombre total de tours'*) passait de 12,46% à 3,21%. Par ailleurs, 75% des participants déclaraient qu'ils voudraient avoir toujours l'affichage visuel dans les services vocaux qu'ils utilisent. Forts de ce résultat, les auteurs ont travaillé sur l'ajout d'un outil dont l'intérêt est de permettre soit la navigation dans les menus soit la recherche directe d'un sous menu. Cet outil a été évalué dans une troisième expérience qui montre que l'outil de recherche est le plus efficace. Il s'agit d'une extension logicielle des possibilités de navigation. On peut noter que ce travail d'ingénierie occulte l'analyse cognitive et comportementale (qui ne rentre pas dans ses objectifs. On ne saurait le reprocher aux auteurs).

Dans l'ensemble, ces expériences ont l'intérêt de montrer que l'utilisation de la modalité visuelle pour présenter certaines indications a un impact positif dans le DHM vocal. Il semble donc intéressant de continuer dans la même direction. C'est l'une des préoccupations dans les travaux sur les modes de communication.

3.2.2 Effets des modes de communication

La notion de *'mode de communication'* renvoie simultanément aux entrées et aux sorties du système. La *'communication écrite'* utilise des entrées au clavier avec des sorties graphiques sur écran. La *'communication vocale'* utilise des entrées produites oralement avec des sorties vocales via haut-parleur. Ces deux modes de communication ont fait l'objet de plusieurs comparaisons en termes d'influence sur le processus collaboratif (Le Bigot, Jamet & Rouet, 2004; Le Bigot, Jamet, Rouet, & Amiel, 2006; Le Bigot et al., 2007; Parush, 2005) ; et également de comparaisons croisées ayant pour but de distinguer les effets spécifiques aux modes de production et de compréhension dans le cadre d'une interaction dialogique (Le Bigot, Rouet & Jamet, 2007).

A Comparaison du mode vocal et du mode écrit

Les résultats montrent que le *'mode écrit'* (clavier et écran) est le plus efficace (Le Bigot, Jamet, & Rouet, 2004). Dans ce mode, le nombre de tours de dialogue était le plus réduit (1,52 à l'écrit *vs.* 3,38 à l'oral) et les dialogues étaient globalement plus courts (61,94 *vs.* 81,37 secondes). De plus, les mesures de *'charge de travail mentale'* (*« mental workload »*) subjective (version modifiée du NASA-TLX) ont donné les évaluations les plus basses en mode écrit (2,71 *vs.* 3,79, sur une échelle en 9 points). Cependant, dans ce même mode, la longueur des tours de dialogue était plus importante (nombre de mots : 6,95 *vs.* 5,28 ; durée par énoncé utilisateur : 32,9 *vs.* 18,2 secondes). Les auteurs en ont conclut que le dialogue vocal induirait une complexité cognitive plus importante en comparaison au dialogue écrit du fait du nombre de mots, du nombre de tours de dialogue et de la charge de travail mentale qui étaient tous plus importants en vocal.

Les articles suivants ont permit de préciser ces résultats. Le processus d'apprentissage a été étudié par Le Bigot, Jamet, Rouet et Amiel (2006). Dans cet article, les auteurs ont montré qu'au cours de 12 dialogues successifs la performance augmentait quel que soit le mode utilisé et qu'un changement de mode soit intervenu ou non à mi-parcours. Ils ont montré qu'indépendamment de l'apprentissage, les pronoms personnels et les articles étaient utilisés plus fréquemment en mode vocal. Là aussi, les auteurs ont discuté les résultats en termes de charge de travail mentale et de ressources. Leur discussion précise que le mode a un *'effet indirect sur l'activité*', via le coût mental, et que la tâche à accomplir est priorisée dans le mode écrit (p. 496).

Ce point a été développé plus explicitement dans un article qui porte plus spécifiquement sur le processus collaboratif (Le Bigot et al., 2007). Les auteurs y ont précisé que les résultats pouvaient être interprétés non seulement au regard du *'principe d'économie cognitive'*, mais aussi en termes d'adaptation. Ils ont indiqué que la présence d'un affichage continu en mode écrit avait permis aux participants de mieux distribuer et gérer leurs ressources. En mode vocal, les indicateurs verbaux (en particulier, les pronoms à la première personne) marquaient l'implication des participants. Ainsi, ces auteurs proposent de considérer que : *« La réutilisation du matériel produit par le système et l'élimination des termes non essentiels pourraient refléter des différences dans le principe d'économie cognitive, qui serait vu comme une fonction du mode de communication. »* (p. 990, traduction libre.)

B Comparaisons croisées

L'article sur le croisement des modes de communication (Le Bigot, Rouet & Jamet, 2007) a permis de préciser certains détails. Dans l'expérience présentée, deux variables étaient manipulées : (1) la modalité d'entrée (clavier *vs.* expression orale) et (2) la modalité de sortie (affichage sur écran *vs.* diffusion de prompts sonores). Ainsi, quatre *'modes de communication'* étaient comparés. Les résultats ont indiqué que la durée du dialogue était affectée surtout par le mode de présentation. Les dialogues étaient plus longs quand le système présentait ses réponses sur la modalité auditive. Le mode de production des

participants a également produit plusieurs effets. Quand les participants commandaient à la voix, leurs énoncés étaient plus longs, le nombre d'erreurs était plus important et le taux de satisfaction tendait à baisser. De plus, une différence est apparue concernant la charge de travail mentale entre la condition entièrement vocale (3,71) et la condition entièrement écrite (2,42). Le nombre de répétitions de la part des utilisateurs variait comme la charge de travail mentale. Ces résultats montrent que les modes de production et de compréhension ont eu des effets différents dans le dialogue. L'interprétation des auteurs est basée sur le coût mental engendré par le mode vocal, qui est associé aux longueurs supplémentaires dans le dialogue (longueur des énoncés vocaux du système, en sortie, et tours supplémentaires pour la correction des erreurs de reconnaissance vocale, en entrée).

A partir de ces résultats, ils ont proposé des conclusions et des extensions pour des recherches futures orientées sur la combinaison des modalités. Cette discussion revendique le rôle prépondérant de la charge de travail mentale de l'utilisateur avant d'aborder les aspects réellement dialogiques. Sur ce point, les recommandations des auteurs portent sur l'utilisation d'entrées vocales et de sorties graphiques. Etant donné les résultats de leur expérience, ils ont supposé que cette combinaison était la plus efficace. Concernant l'utilisation des entrées vocales, une étude sur la charge physique (à l'aide d'EMG) confirme en effet que ce mode réduit l'activité musculaire (Juul-Kristensen, Laursen, Pilegaard, & Jensen, 2004) ; cela pour des raisons liées à l'activité des bras et à l'orientation de la tête. Concernant les sorties, (Le Bigot, Rouet & Jamet, 2007) précisent également qu'il serait utile de réduire la longueur des énoncés, ce qui ferait baisser la charge de travail mentale et permettrait ainsi d'améliorer la collaboration.

C Conclusion sur les modes de communication en DHM

Les commentaires qui peuvent être apportés ici portent avant tout sur les relations de causalité utilisées par les auteurs pour expliquer les résultats. La référence au *principe d'économie cognitive* utilisée ici renvoie à une gestion personnelle de l'effort. En fait, ce point de vue s'oppose à celui de Clark dans deux de ses articles importants (Clark & Wilkes-Gibbs, 1986 ; Clark & Brennan, 1991) où il plaide pour une gestion collaborative de l'effort (*« least collaborative effort »*). Pour cette raison, on peut penser que d'autres causes peuvent être recherchées pour expliquer les différences dans la réutilisation du matériel linguistique ; qui renverraient au processus collectif plutôt qu'à l'effort individuel. L'égocentrisme intrinsèque à la communication, qui a été noté dans le chapitre 1 (*Cf.* 1.4.3), se rapporte à l'interprétation des références (tant en production qu'en compréhension), c'est-à-dire à la construction (individuelle) du sens. Le processus de grounding est, au contraire, intrinsèquement collectif puisqu'il a vocation à coordonner ce sens entre les partenaires. Il est sans doute possible de considérer que la réutilisation du matériel, dans un mode ou dans l'autre, est liée à un jeu de relations interne au processus de grounding et que l'effort mental dépensé pour s'adapter aux contraintes est une résultante de ces relations. Les évaluations subjectives de charge, faites après l'exécution du processus, pourraient être impactées par ce jeu de relation dans le

déroulement de l'activité, plutôt que l'inverse. Cette proposition consisterait, en fait, en un renversement de la chaîne causale proposée par les auteurs qui s'appuient sur la notion de ressources pour expliquer le fonctionnement cognitif.

Les liens entre les notions de ressources et de charge cognitive sont abordés au chapitre 4.

3.2.3 Conclusion sur l'étude des énoncés en DHM

Ces études illustrent les moyens disponibles pour optimiser le processus de communication en DHM. Ces travaux se sont orientés sur l'influence du comportement de l'utilisateur, de façon à ce qu'il produise, autant que possible, les comportements attendus. Différentes techniques ont été isolées pour assurer cette influence en vertu des capacités des systèmes, *i.e.* en tenant compte de la nécessité de restreindre le vocabulaire actif à chaque instant dans les logiciels de reconnaissance vocale. En termes méthodologiques, les études montrent que, dans la mesure où le dialogue implique un processus complexe, il est nécessaire de se reposer sur une diversité d'indicateurs pour proposer un diagnostic sur les situations étudiées (*e.g.* Le Bigot, Rouet & Jamet, 2007).

Dans l'ensemble, les travaux ont été principalement orientés sur une communication auditive dans la mesure où ces systèmes étaient conçus en majorité pour le téléphone classique. Cependant, quelques travaux font appel à l'utilisation de la modalité visuelle et ils offrent des perspectives prometteuses ; et les évolutions actuelles des terminaux et des réseaux permettent d'envisager l'exploitation de ces promesses. Cependant, à ce stade, ces travaux n'ont permis que d'isoler certains effets avantageux. Ils ne permettent pas encore d'intégrer un « comportement multimodal » complet dans les systèmes. Comme l'indiquaient Whittaker et Walker (1991), une telle intégration suppose pour le système d'être capable de faire des choix sur ce qu'il présente. L'acquisition de connaissances supplémentaires est encore nécessaire aujourd'hui pour permettre l'automatisation de ces choix. Notamment, les conditions d'utilisation complémentaire des modalités visuelle et auditive, en sortie, et dans le contexte du dialogue, sont encore mal connues. Une approche psycho-ergonomique de ces questions est nécessaire.

D'autres travaux utilisant la modalité visuelle en complément de la modalité auditive ont été proposés. Ils sont évoqués dans les parties qui suivent au sujet des théories de l'apprentissage multimédia et de la redondance audio-visuelle des informations.

3.3 L'étude de la multimodalité en situation d'apprentissage

Les théories de l'*apprentissage multimédia* (ou *« Multimedia Learning »*, voir Mayer & Coll., 2005) portent sur les différentes manières de combiner les modes de présentation des informations. Dans ce cas, l'utilisateur est vu comme un *'individu apprenant'* (ou *'apprenti'*). L'objectif est de lui permettre le meilleur apprentissage des contenus présentés en se basant

sur le fonctionnement de l'*architecture cognitive*. Comme on l'a vu, ces travaux sont utilisés en ingénierie des systèmes multimodaux (*e.g.* Bachvarova, Van Dijk & Nijholt, 2007) comme travaux de référence en matière de combinaison des modalités. Pour cette raison, il est important de présenter l'approche proposée par ces auteurs.

Les théories les plus importantes du domaine sont la *théorie de la charge cognitive* (Sweller, 1988, 1999, 2005), la *théorie de l'apprentissage multimédia* (Mayer, 2005) et le *modèle intégré de la compréhension de texte et d'image* (Schnotz, 2005). Ce dernier n'est pas abordé dans la thèse dans la mesure où son champ d'application n'entre pas directement dans le cadre d'étude abordé. La *théorie de la charge cognitive* est abordée dans le dernier chapitre théorique (chapitre 4) qui est consacré au problème des ressources de l'utilisateur et de sa charge cognitive.

3.3.1 La théorie de l'apprentissage multimédia

Ces théories relèvent d'un intérêt pour le développement des TICE (*'Technologies de l'Information et de la Communication pour l'Education'*) – voire plus généralement pour la conception des situations de travail, comme le notent Chanquoy, Tricot et Sweller (2007) –. Elles impliquent des réflexions sur la *conception des énoncés*, soit pour des ouvrages éducatifs (livres), soit pour des environnements interactifs. L'expression utilisée par les auteurs est : *« instructional design »* (*'conception instructionnelle'* ou *'conception d'enseignement'*). On pourrait traduire : « *conception de messages à visée instructive* ». Dans la perspective des auteurs, il s'agit de concevoir des formats de présentation qui soient en accord avec les capacités de traitement des individus apprenants. L'objectif est d'utiliser les différents médias disponibles pour concevoir des messages *multimédias* (ou "multimodaux", ces termes étant considérés comme synonymes par Moreno & Mayer, 2007) bénéfiques pour l'apprentissage.

L'objectif de la thèse va dans le même sens, mais il consiste à appliquer cette démarche de conception au cas des systèmes de DHM. Dans cette perspective, la traduction adoptée dans la thèse est : « *conception d'énoncés*[1] *»*. Il s'agit également de concevoir des *'messages multimédias'* qui soient bénéfiques pour le traitement des informations par l'utilisateur, ce qui suppose de prendre en compte l'apprentissage des informations consultées.

Le postulat sous-jacent à cette approche est exprimé par Mayer (2005, page 9) :

> *« La prémisse sous-jacente à l'approche centrée utilisateur est que les conceptions qui sont cohérentes avec le fonctionnement de l'esprit humain sont plus efficaces pour favoriser l'apprentissage que celles qui ne le sont pas. »*

Il s'agit de concevoir des énoncés qui soient en accord avec le fonctionnement de l'esprit humain. Ce fonctionnement est vu sous l'angle des processus de traitement des informations

[1] La traduction anglaise du terme *énoncé* (« *utterance* ») n'est pas utilisée dans ce domaine.

(perceptifs et cognitifs) comme le précise l'article de Schnotz et Kürschner (2007) qui débute par ce paragraphe (traduction libre) :

> *« L'apprentissage et l'instruction ont subi d'importants changements ces dernières années. Les nouvelles technologies permettent la construction d'environnements d'apprentissage qui permettent de présenter l'information électroniquement selon différents formats représentationnels et de manière flexible. Bien que l'aspect technique soit fondamental pour le fonctionnement de ces environnements il n'est pas en soi très intéressant dans une perspective de science psychologique ou instructionnelle. Au lieu de cela, les aspects importants réfèrent aux formats représentationnels et aux processus perceptifs et cognitifs qui apparaissent quand l'apprenti interagit avec les environnements d'apprentissage. Les questions d'importance centrale sont : Que se passe-t-il dans l'esprit de l'apprenti quand du texte parlé ou écrit accompagné d'images fixes ou animées ou de graphiques lui sont présentés ? Comment l'information présentée peut-elle être adaptée aux limites du système cognitif ? »*

Les *limites du système cognitif* renvoient aux modèles de la mémoire et de l'attention. L'accent est mis sur les processus de captage de l'information chez l'individu apprenant (processus perceptifs et cognitifs) et les aspects techniques sont dits *« pas très intéressants »* pour la psychologie. Ainsi, le point de vue est différent de celui de l'*approche intégrée'* adoptée en ingénierie cognitive (prise en compte du *système formé par l'utilisateur et l'outil*. Voir première partie du chapitre 2). Dans le cas de l'apprentissage multimédia, l'approche est *'centrée individu'*. Le système n'est envisagé que sous l'angle des « formats représentationnels » qu'il permet d'utiliser pour présenter l'information.

Le modèle du *fonctionnement cognitif de l'individu apprenant* est basé sur les théories classiques de la mémoire qui focalisent sur les limites des capacités de traitement et sur la notion de *charge cognitive*. L'analyse des environnements d'apprentissage est basée sur les notions d'*information*, de *mode de présentation* et de *modalité perceptive*.

A Postulats théoriques

Dans l'ouvrage de référence du domaine (Mayer & Coll., 2005), Mayer indique que pour étudier l'apprentissage multimédia il souhaite rechercher « *comment l'information pénètre dans le système cognitif »*. Il propose trois hypothèses sur le fonctionnement des voies entrantes (Tableau 3-2, tiré de Mayer, 2005).

Tableau 3-2 : Postulats théoriques de la théorie de l'apprentissage multimédia

Double canaux	L'information peut être traitée visuellement ou auditivement. La *théorie du double codage* (Paivio, 1986) ainsi que les travaux sur la *structure de la mémoire à court terme* (Penney, 1989) indiquent que deux types de codage sont utilisés par les individus mis dans une situation d'apprentissage. Ils indiquent également qu'un encodage utilisant les deux systèmes à la fois permet une meilleure mémorisation car l'apprentissage est plus profond.
Capacité limitée	La mémoire est envisagée sous l'angle du modèle de Baddeley (Baddeley, 2004; Baddeley & Hitch, 1974). Un rôle important est accordé au *traitement exécutif central* (*« central executive »*) pour le

	contrôle de l'allocation des ressources cognitives au cours de l'activité. L'accent est mis sur les limitations des capacités de traitement lors de l'apprentissage.
Traitement actif des informations	L'individu apprenant s'engage activement dans la construction d'une représentation mentale. La *théorie des schémas* (Johnson-Laird, 1983) est généralement utilisée pour évoquer ces représentations mentales stockées en mémoire à long terme. La construction de ces schémas implique des processus d'*attention*, de *sélection*, d'*organisation* et d'*intégration* dont la complexité dépend de la structure des connaissances à intégrer. Pour la conception des énoncés, cela implique (1) de veiller à la cohérence de la structure présentée et (2) de veiller au guidage de l'individu apprenant (Moreno & Mayer, 2005).

Ces postulats théoriques permettent de proposer des hypothèses sur les processus actifs lors de l'apprentissage et sur les performances qui en découlent dans des *tests de rappel* et dans des *tests des connaissances acquises* (résolution de problème, transfert des connaissances sur des problèmes proches). Dans la mesure où l'objectif est de faciliter l'apprentissage des contenus, toute l'attention est portée sur les processus individuels que met en jeu une personne mise dans une situation d'apprentissage.

B Analyse des environnements d'apprentissage

Moreno et Mayer (2007) indiquent que leurs études passées n'ont pas pris en compte l'interactivité des environnements d'apprentissage et qu'ils souhaitent remédier à ce manque.

Ils définissent :

- Les *'environnements d'apprentissage multimodaux'* comme : *« les environnements d'apprentissage qui utilisent deux modes différents pour présenter l'information »*.
- Les *'environnements d'apprentissage multimodaux interactifs'* : *« sont* ceux *dans lesquels ce qui arrive dépend des actions de l'apprenti »*.

L'analyse des environnements d'apprentissage proposée par ces auteurs repose sur le type de codage des contenus et sur les canaux utilisés. Les auteurs opposent la notion de *'mode de présentation'*, qui renvoie au *code* utilisé (*'verbal'* ou *'non verbal'*), à celle de *'modalité perceptive'* (*'auditive'* ou *'visuelle'*). A partir de ces éléments ils définissent un *'principe de modalité'* qui stipule que : *« les environnements d'apprentissage les plus efficaces sont ceux qui combinent représentation verbale et non-verbale des connaissances en mixant les modalités pour la présentation »* (Moreno & Mayer, 2007, p. 310). Ainsi, ils considèrent, en se basant sur les postulats théoriques présentés précédemment, que la présentation strictement visuelle des matériaux (verbaux et non verbaux) surcharge les capacités cognitives pendant l'apprentissage.

Dans les environnements interactifs, la communication est dite « multidirectionnelle » (Markus, 1987). Dans ce contexte : *« les buts des actions des participants doivent être de favoriser l'apprentissage »* (Moreno & Mayer, 2007, p.311). Aucun autre type de but n'est mentionné. L'interactivité est ensuite définie en fonction du type d'action permis à l'apprenti : (1) dialogue, (2) contrôle, (3) manipulation, (4) recherche, (5) navigation. Pour les auteurs, un continuum

peut être définit selon que l'environnement est plus ou moins interactif (*« continuum d'interactivité »*). Les auteurs considèrent que, dans la mesure où l'individu est actif, il met en place des processus cognitifs et comportementaux qui peuvent entrer en concurrence. L'interactivité peut créer une *'charge cognitive inutile excessive'* (*« excessive extraneous load »*) qui perturbe l'apprentissage. Ainsi, en termes explicatifs, les auteurs s'en remettent à la *théorie de la charge cognitive* (Sweller, 1999; 2005).

Finalement, les notions utilisées pour analyser les environnements d'apprentissage sont celles d'*information'*, de *'code'*, de *'modalité'* et d'*interactivité'*. Cette analyse permet de prendre en compte les moyens dont dispose *le système* pour exprimer les informations que l'apprenti doit acquérir et les moyens dont dispose *l'apprenti* pour percevoir et construire les connaissances qui intègrent ces informations.

C La synthèse théorique de Mayer

Mayer et Moreno (Mayer, 2005; Mayer & Moreno, 2003) ont d'abord proposé une *'théorie cognitive de l'apprentissage multimédia'* (*« cognitive theory of multimedia learning »*). Celle-ci était basée sur des études faites avec du matériel non interactif. Elle a été mise a jour en une *'théorie cognitive-affective de l'apprentissage à partir de médias'* (*« cognitive-affective theory of learning with media »*, Moreno & Mayer, 2007), qui intègre également les aspects interactifs, selon les auteurs.

Sept hypothèses sont à la base de cette théorie : (1) l'existence de canaux séparés, (2) la limitation des traitements possibles à un instant donné, (3) l'apprentissage dépend de l'effort conscient de l'apprenti, (4) la mémoire à long terme est une structure dynamique qui intègre des connaissances épisodiques et sémantiques, (5) les facteurs motivationnels sont des médiateurs de l'apprentissage, positivement ou négativement, (6) les facteurs métacognitifs sont des médiateurs de l'apprentissage par régulation des processus cognitifs et affectifs et (7) les connaissances et habiletés préalables peuvent affecter l'apprentissage.

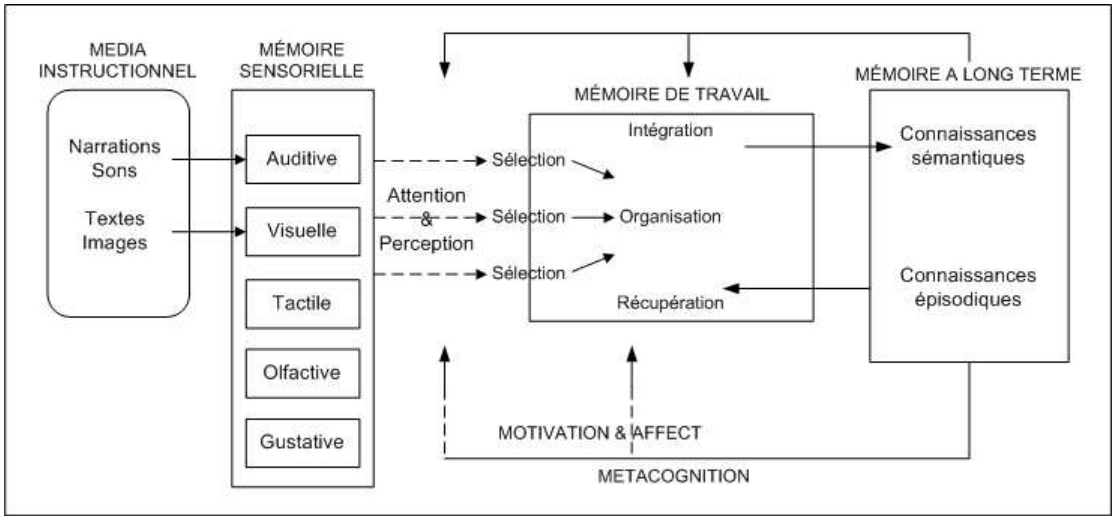


Figure 3-1 : Théorie cognitive-affective de l'apprentissage à partir de médias

Le schéma proposé est présenté dans la Figure 3-1. Cette version améliorée ne modifie pas la philosophie du modèle proposé préalablement (voir Mayer, 2005). L'information circule en démarrant du *'média instructionnel'*, à gauche. Elle est perçue grâce aux mécanismes attentionnels, sélectionnée, intégrée et organisée pour être stockée en mémoire à long terme, à droite, où elle peut être récupérée. Les flèches situées en haut et en bas du schéma indiquent l'effet régulateur des processus affectifs et métacognitifs lors de l'apprentissage. Mais le processus d'apprentissage est vu comme une circulation des informations de gauche à droite. Aux différentes étapes de ce cheminement, la capacité de traitement peut être dépassée de sorte qu'une *'surcharge cognitive'* apparaît.

Après avoir présenté ce schéma, les auteurs introduisent les notions de *'demande de traitement'* (*« processing demands »*) et de *'capacité de traitement'* (*« processing capacity »*). A partir de cette distinction, ils définissent des processus qui peuvent être essentiels ou gênants pour l'apprentissage, car ils facilitent ou compliquent le traitement des informations. Ces processus renvoient à la *théorie de la charge cognitive*. Ils sont abordés au chapitre suivant.

3.3.2 Principes de conception des énoncés pour l'apprentissage

Bien que ces théories soient conçues dans le but de promouvoir la conception des environnements interactifs dédiés à l'apprentissage, les auteurs ne développent pas explicitement une *méthode de conception des énoncés*. Lors de la présentation d'une expérience, différentes combinaisons des modes de présentation sont présentées dans l'introduction de l'article, puis elles sont converties sous forme de conditions expérimentales. Chaque participant est confronté à une de ces combinaisons pour créer autant de groupes dont les performances d'apprentissage sont comparées. Comme cela a été noté par Schnotz et Kürschner (2007), la réalisation technique des combinaisons proposées n'est pas jugée pertinente pour la recherche en psychologie.

Les principes de combinaison sont basés sur les éléments dégagés dans l'analyse des environnements d'apprentissage :

1. ***Catégorisation*** : Les informations sont catégorisées en fonction du code qui permet de les exprimer. Elles peuvent être soit *'verbales'*, soit *'non verbales'* (graphiques, imagées ou animées) ;
2. ***Attribution d'un mode de présentation*** : Un mode de présentation est attribué à chacun de ces types d'information. L'information *verbale* peut être présentée visuellement, en *mode écrit*, ou auditivement, en *mode vocal*. L'information *non verbale* est le plus souvent présentée visuellement, en *mode graphique*. Certaines informations *non verbales* peuvent être présentées dans un *mode sonore*.

Sur cette base, des hypothèses sont proposées sur les processus cognitifs qui peuvent être mis en place par l'individu apprenant confronté à l'une ou l'autre des combinaisons.

L'hypothèse principale consiste à prédire que l'on apprend plus facilement à partir de mots illustrés d'images, c'est-à-dire en utilisant à la fois le *code verbal* et le *code imagé*, qu'à partir de mots seuls (Mayer, 2005).

Cette méthode permet de déduire des *'principes cognitifs pour la conception des énoncés'* (*« cognitive principles of instructional design »*). Les *effets* obtenus dans les expériences sont convertis sous forme de *principes de conception*. Les effets les plus fréquemment évoqués (*Cf.* Tableau 3-3) sont l'*effet de modalité* et l'*effet de redondance* – qui donnent lieu aux principes du même nom –, les *principes de contiguïté spatiale* et *temporelle* et le *principe de cohérence*. Les travaux les plus récents ont également permis d'obtenir des effets liés aux environnements interactifs tels que le *principe d'activité guidée*, le *principe de réflexion*, le *principe de contrôle pas-à-pas* (*« pacing »*), etc. L'objectif du Tableau 3-3 est de présenter succinctement les différents effets plutôt que de lister exhaustivement les effets connus.

Mayer et Moreno (2007) remarquent qu'il est nécessaire de vérifier si les effets qui ont été obtenus dans des environnements non interactifs s'appliquent dans le cas des environnements interactifs. Ainsi, ils reconnaissent l'importance du contexte dans les effets qui apparaissent et ils indiquent qu'il existe probablement une dépendance contextuelle qui pourrait expliquer les effets déjà connus. En termes de conception, cela implique qu'il est nécessaire de relativiser les principes de conception les uns par rapport aux autres.

Tableau 3-3 : Principes de conception pour l'apprentissage multimédia

Principe	Effet
Modalité	L'*effet de modalité* indique que l'apprentissage est favorisé si l'information textuelle qui accompagne une information visuelle (graphique ou imagée) est présentée sur la modalité auditive plutôt que sur la modalité visuelle. Cet effet a été très largement observé. Ginns (2005) a présenté une méta-analyse de 43 études qui montre la robustesse de cet effet. Cet auteur indique que l'interactivité de l'environnement et le contrôle pas-à-pas du défilement des informations (*« pacing »*) ont des effets modérateurs importants sur l'effet de modalité. Il est donc important de prendre en compte l'interactivité de l'environnement d'apprentissage.
Redondance	L'*effet de redondance* est un effet négatif lié à la présentation inutilement redondante des informations, qui surcharge la mémoire de travail (Diao & Sweller, 2007). C'est le cas si les mêmes informations sont présentées à la fois sur la modalité visuelle et sur la modalité auditive (Sweller & Chandler, 1994). Cet effet est également présent si on présente à des experts des informations qu'ils connaissent déjà (Kalyuga, Ayres, Chandler, & Sweller, 2003). Ses connaissances incitent l'expert à vérifier leur cohérence avec celles qui sont présentées, ce qui produit un effet négatif sur l'apprentissage. Ces résultats sont interprétés comme des surcharges de la mémoire de travail. On parle aussi d'*effet de cohérence* par opposition à l'*effet de redondance*. Mais la redondance audio-visuelle peut également produire un effet positif dans certains cas (Moreno & Mayer, 2002). Dans ce cas, les auteurs interprètent que cet effet est lié à un accroissement de l'espace disponible en mémoire de travail. Le Bohec et Jamet (2005) remarquent la polysémie du terme, qui permet ces deux interprétations. Ces auteurs concluent à la nécessité d'adapter la présentation à chaque utilisateur, en fonction de son niveau d'expertise et de ses demandes au cours de l'interaction.

Contiguïté spatiale / temporelle	Les *effets de contiguïté spatiale* et de *contiguïté temporelle* sont des effets positifs pour l'apprentissage liés, respectivement, à la structuration spatiale et à la synchronisation temporelle des éléments présentés. Le fait d'intégrer les informations textuelles au graphique/schéma est favorable à l'apprentissage , tout comme le fait de présenter les messages auditifs pendant la présentation visuelle du graphique/schéma (Mayer & Moreno, 1998). L'interprétation de ces résultats repose sur le partage de l'attention. Dans la mesure où les ressources cognitives sont limitées, le partage de l'attention entre plusieurs sources d'information est coûteux et défavorable pour l'apprentissage.
Activité guidée	Par ailleurs, la théorie *cognitive-affective de l'apprentissage à partir de médias* (Moreno & Mayer, 2007) permet d'introduire une série de principes liés au guidage de l'apprenti dans l'activité. Cinq principes sont évoqués : (1) *Principe d'activité guidée* : mieux vaut guider l'utilisateur que lui donner des explications sans le guider ; (2) *Principe de réflexion* : mieux vaut inciter l'apprenti à réfléchir aux réponses correctes ; (3) *Principe de feedback* : mieux vaut des feedbacks explicatifs que des feedbacks correctifs ; (4) *Principe de contrôle pas-à-pas* : mieux vaut permettre à l'apprenti de contrôler le défilement des informations ; (5) *Principe d'amorçage des connaissances* : une activité préalable d'activation des connaissances permet d'améliorer l'apprentissage.

3.3.3 Conclusion

Finalement, les théories de l'apprentissage multimédia sont basées sur une approche centrée sur l'individu. L'approche se base sur une synthèse des modèles classiques de la mémoire de travail pour dégager des principes de conception applicables aux environnements interactifs. Ces principes de conception sont basés sur les distinctions (1) entre code *verbal* et *non verbal* et (2) entre modalité *visuelle* et *auditive*, qui permettent de proposer différents modes de présentation des informations. Les résultats qui en découlent peuvent être utilisés comme guides de conception par les praticiens, sous forme de recommandations.

Les effets mis en évidence, ainsi que la méthode de conception des énoncés qui leur correspond, sont analysés relativement à la *tâche d'apprentissage* assignée aux participants des expériences. Dans l'analyse proposée, cette tâche d'apprentissage est la seule considérée. L'existence de tâches concurrentes qui apparaîtraient au cours de l'activité d'apprentissage n'est pas envisagée dans les articles cités. Pourtant, les travaux sur le dialogue indiquent que cette activité peut être décomposée au moins en deux familles de *fonctions* : (1) les *fonctions d'information*, qui correspondent à la tâche d'apprentissage et (2) les *fonctions de contrôle*, qui correspondent à la tâche de gestion de l'interaction. Cette analyse peut facilement être étendue à tous les types d'activité finalisée avec un

environnement interactif, ou même non interactif[1]. Notamment, elle peut être appliquée au cas de l'utilisation d'un environnement d'apprentissage. Une critique allant dans ce sens a été proposée par des chercheurs du domaine (Gerjets & Scheiter, 2003). Ces auteurs indiquent que l'appariement (« *mapping* ») entre les *modes de présentation* et les *patterns de charge cognitive* qui en résulte est trop direct. Pour eux, les buts de l'enseignant et les activités de l'apprenti jouent des rôles modérateurs importants qui modifient la relation entre mode de présentation et charge cognitive. On a vu que Ginns (2005) faisait des remarques allant dans le même sens. La relation de causalité entre principes de conception, charge cognitive et apprentissage est modérée par ces auteurs, mais elle n'est pas remise en cause.

La Figure 3-2 est une interprétation du point de vue présenté par Mayer et ses collaborateurs dans le même système de représentation que celui adopté dans la Figure 1-1 (page 11) pour l'interprétation du point de vue d'Austin (1962). Dans la Figure 3-2, le système multimédia permet de présenter des informations verbales et non verbales qui sont perçues par l'apprenti grâce aux deux modalités (visuelle et auditive) puis intégrées à ses connaissances. Par rapport à la Figure 1-1, le locuteur a été remplacé par le système multimédia pour indiquer que ce sont les effets des énoncés du système qui sont étudiés.

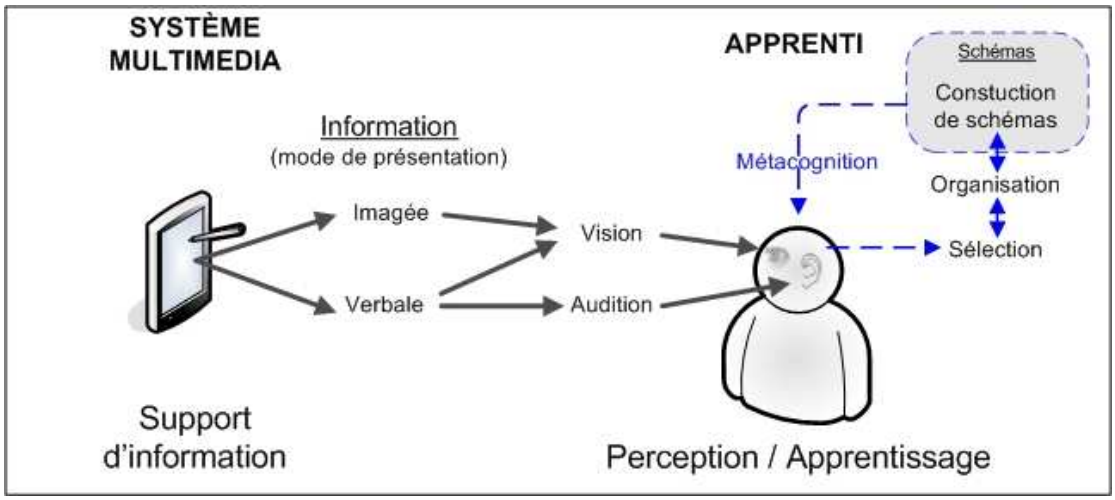


Figure 3-2 : Interprétation de la théorie de l'apprentissage multimédia

La différence importante tient à la *disparition de la dimension illocutoire des actes*. Le système produit des actes d'énonciation à gauche du schéma. Ces actes produisent des effets sur l'apprenti, à droite. Mais aucune relation conventionnelle ne permet de qualifier la situation dans laquelle ces individus sont engagés. L'information est vue comme une entité qui circule de gauche à droite. Même lorsque les auteurs prennent en compte les connaissances préalables de l'individu apprenant, il ne s'agit pas de montrer comment cet individu utilise ses connaissances dans l'interaction. Il s'agit seulement de mesurer la différence entre le niveau de ses connaissances avant et après l'interaction. De ce fait, par définition, cette théorie considère que le système est source de connaissance et l'apprenti est

[1] Un livre est un environnement non interactif. Mais même dans ce cas, il faut, par exemple, le maintenir ouvert. C'est une fonction de contrôle.

source d'ignorance. Ce positionnement n'est jamais remis en question. De plus, les effets produits sur l'apprenti consistent à intégrer les informations qui lui sont présentées. Dans la perspective d'Austin (1962), seuls les *effets perlocutoires* sont envisagés ici. Austin ajoutait les *suites perlocutoires* qui correspondent aux réponses comportementales (y compris verbales) de l'apprenti dans la situation. Schegloff (2006), tout comme Clark (2004), mentionnent ces actes comme des *obligations de réponse*, ce qui souligne leur nature conventionnelle. A ce stade, ces éléments ne sont pas pris en compte dans les théories de l'apprentissage multimédia. Dans les termes d'Austin, on pourrait dire que ces théories relèvent de l'*illusion descriptive*.

D'autres critiques sont également proposées au chapitre suivant, autour de la notion de charge cognitive. Ces critiques dans leur ensemble portent sur les aspects fondamentaux des théories, c'est-à-dire sur le principe explicatif adopté. Elles ne remettent pas en cause l'intérêt pratique des travaux réalisés et des recommandations qui en découlent.

3.4 Le problème de la redondance des informations

Le problème de la redondance des informations est commun à de nombreux travaux autour des communications. Ce problème se pose dans la mesure où, quand des informations ont été identifiées, le choix doit être fait de les présenter sur une modalité unique (auditive ou visuelle) ou sur les deux à la fois. Une première approche intuitive de ce choix, tend à admettre que la présentation simultanée sur les deux modes augmente l'acuité perceptive, si ce n'est la capacité disponible en mémoire de travail. Mais les auteurs indiquent généralement qu'à leur surprise cela ne correspond pas aux résultats qu'ils observent (*e.g.* Helleberg & Wickens, 2003 ; Sweller & Chandler, 1994).

3.4.1 Utilisation conjointe du mode visuel et du mode auditif

La proposition d'utiliser les différentes modalités pour présenter l'information a été testée dans divers contextes d'activité. C'était le cas de Chandler et Sweller dans un contexte d'apprentissage (Chandler & Sweller, 1991; Sweller & Chandler, 1994). D'autres auteurs ont testé, par exemple, l'impact sur des tâches de manipulation (Vitense, Jacko & Emery, 2003) puis plus particulièrement pour des personnes malvoyantes (Jacko et al., 2005).

Dans le cadre d'une tâche d'apprentissage, la redondance audio-visuelle des informations a le plus souvent un effet négatif. Par exemple, Jamet et Le Bohec (2007) ont présenté les commentaires d'un diagramme soit *auditivement*, soit sur un mode *redondant séquentiel* (apparition progressive des phrases au cours de la diffusion des énoncés auditif), soit sur un mode *redondant statique* (affichage de la totalité du texte dès le début de la diffusion de l'énoncé auditif). Les résultats ont indiqué que la présentation visuelle n'aidait pas les participants à mémoriser les contenus. Au contraire, la performance d'apprentissage tendait à se dégrader dans les conditions avec redondance. Les auteurs appuient leurs explications

sur la *théorie de l'apprentissage multimédia* et renvoient à un phénomène de surcharge du canal perceptif visuel au cours de l'activité d'apprentissage.

En dehors des tâches d'apprentissage, une série d'expérimentations a notamment été réalisée par Wickens et ses collaborateurs dans des contextes tels que la *navigation aérienne*, la *surveillance de patients* ou la *conduite automobile* (Helleberg & Wickens, 2003; Seagull, Wickens & Loeb, 2001; Wickens, 2000; Wickens, Goh, Helleberg, & Talleur, 2002). Cet auteur présente également une méta-analyse de 18 études sur l'emploi des modalités auditive et visuelle pour l'information du conducteur en conduite automobile (Wickens & Seppelt, 2002). Dans les conclusions de cette méta-analyse, les auteurs indiquent que la présentation visuelle des informations peut perturber la tâche principale de conduite du véhicule sous certaines conditions. Les études indiquent que dans ce type d'activité la présentation redondante des informations conduit aux mêmes performances que la présentation auditive, alors que la présentation visuelle seule donne lieu à des résultats variables selon les études. Mais dans certaines études (*e.g.* Helleberg et Wickens, 2003), la redondance peut également produire un effet négatif. Finalement, comme c'était le cas dans les travaux issus des théories de l'apprentissage, les auteurs renvoient à des variables modératrices qui atténuent la relation de causalité directe entre mode de présentation utilisé et performance à la tâche. Wickens et Seppelt (2002) identifient cinq variables importantes : (1) la pertinence des informations présentées pour la tâche de conduite, (2) la localisation de l'affichage visuel, (3) la charge de travail associée au déroulement actuel de la tâche de conduite, (4) la complexité des informations présentées et (5) le niveau de redondance entre les informations.

3.4.2 L'effet de préemption

L'étude d'Helleberg et Wickens (2003) permet à ces auteurs d'introduire l'*effet de préemption*, qui met bien en évidence le type d'effet lié à l'emploi des modes de présentation dans des contextes multitâches. Par ailleurs, cet article fournit un bon exemple du type de travaux réalisés autour de la redondance (comparaison simple de trois modes de présentation : *auditif*, *visuel* et *redondant*).

L'étude d'Helleberg et Wickens (2003) porte sur les modes utilisés pour présenter les informations d'une interface de communication (*« data-link »*) dans un contexte de pilotage aérien. La tâche était réalisée par des pilotes entraînés. Dans ce contexte, la hiérarchie des tâches est : (1) pilotage de l'avion (*« aviate »*), (2) navigation aérienne (*« navigate »*) et (3) communication avec la tour de contrôle (*« communicate »*). Les deux tâches principales sont assurées à partir d'information visuelles. Les informations dont la présentation était manipulée portaient sur la troisième de ces tâches (la tâche de communication avec l'interface *« data-link »*). Les auteurs souhaitaient vérifier si le mécanisme attentionnel le plus actif est un mécanisme de *compétition pour les ressources* (*« resource competition »*) ou un mécanisme de *préemption* dans lequel la présentation des informations interrompt

(*« preempt »*) l'activité en cours (dans le cockpit en l'occurrence). Cet effet serait lié à la nature intrusive de la parole et à une caractéristique d'alerte associée aux informations auditives (Posner, Nissen & Klein, 1976). D'après les auteurs, certains travaux antérieurs avaient pu montrer un effet bénéfique de la présentation redondante des informations, surtout dans le contexte automobile ; mais aucune étude n'avait encore testé de façon systématique la présentation soit *auditive*, soit *visuelle*, soit *redondante* des informations dans un contexte multitâches et avec toutes les mesures nécessaires pour en tirer des conclusions fortes au sujet de la distinction entre *effet de préemption* et *compétition pour les ressources* (Helleberg & Wickens, 2003, p. 192).

Les résultats ont montré que la version visuelle donnait lieu aux meilleurs résultats pour l'ensemble des tâches. La version auditive a eu tendance à faire décroître l'attention portée aux informations nécessaires à la tâche de pilotage (*« aviate »*) et elle a conduit à la pire performance en ce qui concerne la détection du trafic (*« navigate »*). Dans cette condition (auditive), les pilotes devaient prendre des notes pour certaines informations de façon à éviter de les mémoriser, mais cela ne compensait pas totalement la présence visuelle de ces informations puisque plus d'attention était requise lors de la prise de note elle-même. La version redondante a donné lieu à des résultats intermédiaires. Dans cette condition, les auteurs pensaient obtenir le meilleur des deux alternatives (*« the best of both worlds »*, p. 208), mais le résultat obtenu leur laisse supposer que les informations auditives ont eu un *effet de préemption* qui a interrompu les autres activités. Pour cette raison, ils ont conclu que l'affichage visuel des informations de communication (*« data-link »*) est préférable dans un contexte de pilotage aérien.

3.4.3 Conclusion

Ces études ont l'avantage d'utiliser des méthodologies écologiques, qui fournissent des résultats valides dans un contexte d'activité multitâches. Mais le problème de la redondance des informations se pose toujours dans les mêmes termes. L'information à présenter est catégorisée en fonction de son code et un mode de présentation simple (auditif ou visuel) ou redondant lui est attribué. Cette analyse donne lieu à des résultats que les auteurs ont du mal à expliquer (comme ils le disent eux-mêmes, Helleberg & Wickens, 2003, p. 208). Malgré l'intérêt de la mise en évidence de l'effet de préemption, des efforts supplémentaires sont encore nécessaires pour fournir l'analyse de ces effets dans le déroulement de l'activité de communication.

3.5 Synthèse

Cette présentation a d'abord porté sur les études empiriques réalisées dans le cadre du DHM. Dans ce domaine, les premiers travaux ont principalement pointé la représentation que l'utilisateur avait du système. Ils ont progressivement permis de mettre en évidence des

techniques utilisables comme principes de conception pour les services commerciaux. Les travaux, plus récents, sur les modes de communication ont permis d'étudier les modifications qui interviennent dans le processus de grounding selon qu'un mode ou l'autre est utilisé pour communiquer, en entrée ou en sortie. Pour présenter les informations, en sortie, la présentation visuelle semble la plus efficace.

Les analyses proposées en psychologie cognitive reposent sur un raisonnement en deux étapes qui permet (1) de catégoriser les informations à présenter en fonction de leur type (qui correspond ici au code, *i.e.* verbal, graphique ou imagé) et (2) d'attribuer un mode de communication à chaque type. Ce raisonnement peut être implémenté dans des systèmes sous forme d'algorithmes (*e.g.* Bachvarova, Van Dijk & Nijholt, 2007). Les explications fournies par les psychologues renvoient aux limites des capacités de traitement des utilisateurs et à la notion de charge (*'charge de travail mentale'* : Le Bigot et al., 2007 ; *'charge cognitive'* : Moreno & Mayer, 2007 ; Sweller, 2005). Selon ces auteurs, les modes de présentation utilisés engendreraient différents types de coût cognitif qu'il faut déterminer pour prédire la performance résultante. Ainsi, les surcoûts cognitifs qui apparaissent parfois suite à la présentation d'informations conduiraient à des modifications, dans un cas, du processus collaboratif, et dans l'autre, du processus d'apprentissage. Ce positionnement qui accorde un statut explicatif au coût de l'activité est très largement répandu en psychologie des apprentissages, même chez des auteurs qui proposent un point de vue critique (*e.g.* Gerjets & Scheiter, 2003) ; et il est largement admis en psychologie des communications (Clark & Brennan, 1991 ; Le Bigot et al., 2007 ; Walker et al., 2004) sans apparente contradiction avec l'importance accordée au processus collaboratif (Clark & Wilkes-Gibbs, 1986). Cependant, d'autres auteurs (Wickens & coll.) tendent à opposer ces explications basées sur les ressources cognitives de l'individu à d'autres explications basées sur le rôle primaire du processus interactif (*effet de préemption*). Pour approfondir ce sujet, la question de la charge cognitive de l'utilisateur est abordée plus spécifiquement dans un dernier chapitre théorique (chapitre 4).

L'enjeu des études est d'isoler les relations de causalité qui opèrent dans des processus complexes tels que le dialogue ou l'utilisation d'environnement multitâche (pilotage aérien, conduite automobile). Ces relations sont analysées à partir de la relation entre type d'information et mode de présentation. Mais, là aussi, l'information considérée est catégorisée sur la base du code qui permet de l'exprimer (verbal ou imagé), ce qui n'est pas une catégorisation très fine. L'analyse expérimentale qui en découle porte sur la redondance de ces informations et relève d'une vision additive des capacités des canaux perceptifs. Pour aller plus loin, la prise en compte d'une *typologie des actions* telle que celle proposée par Nievergelt et Weydert (1980, notions de *'site'*, *'mode'* et *'trace'*, *Cf.* chapitre 2) pourrait permettre d'étudier plus finement cette *relation type-mode*.

Chapitre 4 La *'Charge Cognitive'* de l'utilisateur

On a vu au cours du chapitre 3 que la notion de charge occupe un statut privilégié dans la littérature dans la mesure où la plupart des auteurs y font référence pour expliquer les résultats de leurs expériences. Mais la plupart des auteurs notent aussi que la *'charge cognitive'* est mal définie. On peut alors se demander sur quoi repose le statut explicatif accordé à ce *'concept hypothétique'* (*« hypothetical construct »*) qui acquiert par là le statut d'une *'variable active'* (*« intervening variable »*), contrairement à la distinction classique de ces notions (Mc Corquodale & Meehl, 1948). Ainsi, ce chapitre permet de revenir sur les fondements théoriques et méthodologiques de la notion de charge dans le but de comprendre quel rôle causal elle exercerait dans l'exécution d'une tâche.

Dans cette revue, il ne s'agit pas de faire une présentation exhaustive des travaux qui font référence à la *'charge mentale'*[1] puisque cette notion est présente dans de (trop) nombreux domaines. L'objectif est de rappeler les origines de la notion et le statut explicatif des modèles correspondant dans le fonctionnement cognitif.

4.1 La notion de 'charge cognitive'

La notion de charge s'est développée en psychologie surtout à partir des années 1960, soit à l'époque de l'épanouissement à la fois de la psychologie cognitive et de l'ergonomie. Elle a eu une résonance importante dans la discipline. D'après Richard (1996) : *« cette notion a transité dans plusieurs secteurs de la psychologie : la psychologie expérimentale, la psychologie du travail, la psychologie du développement, la psychologie cognitive ».* La *'psychologie ergonomique'*, plus récente (voir Hoc & Darses, 2007), peut être ajoutée à cette liste car elle y consacre l'un de ses groupes de travail (*« Aspects intensifs »*).

D'après Tsang et Wilson (2004), le nombre de publications sur la notion de charge a connu un succès grandissant dans les années 1970 qui s'est accru dans les années 1980 et à eu tendance à décroître par la suite. Un consensus a émergé dans une série de rencontres qui ont eu lieu dans la seconde moitié des années 1970 (notamment le *NATO Workshop*, 1977 et le *Symposium sur la charge mentale* au *Congrès de Psychologie de Paris*, 1978) (voir, Leplat & Welford, 1978; Moray, 1979). D'après Tsang et Wilson (2004) les objectifs étaient (1) d'arriver à un consensus en ce qui concerne la définition de la charge mentale, (2) d'intégrer

[1] Le terme *'charge mentale'* (*« mental load »*) est théoriquement neutre. Il est utilisé ici pour renvoyer indifféremment à toutes les approches théoriques de la notion. Les termes de *'charge de travail'* (*« workload »*) et de 'charge cognitive' (*« cognitive load »*) sont plus spécifiques. Ils sont expliqués plus loin dans cette partie.

les connaissances existantes pour formuler une théorie de la charge mentale et de son évaluation en contexte applicatif et (3) d'identifier les relations entre les différentes mesures. Mais les auteurs notent que des progrès ont été faits plus particulièrement dans le sens du développement des techniques de mesures et que le consensus théorique est difficile dans la mesure où il supposerait une théorie complète de la performance humaine basée sur la notion de charge mentale.

Cette partie permet d'abord de proposer les définitions des notions de *charge cognitive*, de *ressources* et de *capacité cognitive*. Les paradigmes et les modèles les plus représentatifs sont ensuite présentés plus longuement, avant d'aborder le problème de la mesure de la charge cognitive d'un individu lors de la réalisation d'une tâche.

4.1.1 Définitions

- ***Notion de « charge cognitive »***

Le *concept hypothétique 'charge cognitive'* est supposé indépendant du contexte d'usage. Il s'agit de la quantité d'énergie nécessaire à un individu pour exécuter une tâche. Barrouillet (1996) précise ce point de vue :

> *« La charge cognitive liée à la réalisation d'une tâche donnée peut être considérée comme le niveau d'effort mental (la quantité de ressources) requis par la planification et la mise en œuvre d'une procédure de résolution donnée chez un sujet dont le niveau de développement et d'expertise dans le domaine concerné sont fixés. »* (Page 321)

D'après cette définition, les *« phénomènes baptisés de charge mentale »* (expression de Theureau, 2002) apparaissent dans la relation entre *l'individu* et *la tâche*. Dès lors, deux points de vue différents ont eu tendance à coexister dans la littérature, selon que les auteurs privilégiaient l'analyse des capacités de l'individu ou l'analyse des contraintes liées à la tâche. Vidulich (2003) évoque cette distinction en nommant (1) *'théories de la boîte noire'* (*« black box theories »*) celles qui portent sur le système cognitif de l'individu et qui tentent d'y associer entrées et sorties et (2) *'théories structurales'* celles qui décrivent à un niveau logique les opérations nécessaires à l'accomplissement de la tâche. Ces deux points de vue ont donné lieu à des travaux de natures assez différentes, qui opposent (schématiquement) des travaux sur les ressources et les processus cognitifs à des travaux portant sur la gestion de l'activité et des surcharges épisodiques. Cette distinction se retrouvera dans la présentation des deux questionnaires d'évaluation subjective de la charge (*'Workload Profile'* et *'NASA-TLX'*).

- ***Notions de « ressources » et de « capacité cognitive »***

Dans sa revue théorique sur le sujet, Barrouillet (1996 : « *Ressources, capacités cognitives et mémoire de travail »*) rappelle les définitions essentielles et les postulats nécessaires au fondement de la notion de charge cognitive. Il distingue d'abord les notions de *'ressources'* et de *'capacité cognitive'* de celle de *'charge cognitive'* :

> Les *« ressources »* désignent *« l'énergie mentale disponible pour un individu particulier à un instant donné pour une classe particulière de traitement ».*
>
> La *« capacité cognitive »* renvoie *« à la quantité maximale de ressources que peut mobiliser un individu.*

Mais après avoir proposé ces définitions pour justifier l'emploi de la notion de charge cognitive, l'auteur note que cette dernière dépend au moins de trois autres types de facteurs : (1) le *niveau d'expertise* du sujet, (2) son *niveau de développement* et (3) la *stratégie adoptée* pour résoudre la tâche. Etant donné ces facteurs, il est impossible de prédire quelle quantité de ressources sera mobilisée par un individu pour réaliser une tâche. De plus, la notion de charge cognitive dépend du modèle de ressources sous-jacent, dont la nature peut varier. Ces notions semblent donc problématiques. La critique de Navon (1984) va même plus loin et indique que les phénomènes attentionnels expérimentaux seraient explicables par d'autres facteurs qu'une quantité de ressources limitée.

4.1.2 Paradigmes et modèles de la charge cognitive

Pour clarifier la notion de charge cognitive et tenter de répondre à ces critiques Barrouillet (1996) présente deux métaphores fréquemment utilisées pour décrire les modèles de ressources. Il rappelle que la notion de *'ressource'* est elle-même une métaphore qui permet de décrire le fonctionnement cognitif. Cela induit que la métaphore adoptée est représentative de l'approche conceptuelle choisie par un auteur. D'après Barrouillet :

(1) La *« métaphore spatio-temporelle »* est basée sur l'espace de stockage disponible en mémoire de travail. Elle renvoie à deux fonctions : (1) stockage et (2) traitement des informations, auxquelles correspondent deux types de limitations : (1) concernant l'espace de stockage et (2) concernant le temps de traitement ;

(2) La *« métaphore énergétique »* est inspirée des études en neuropsychologie sur l'attention. Dans cette approche, les ressources sont vues comme la partie active de la mémoire à long terme. Chaque élément de la mémoire est caractérisé par un certain seuil d'activation, selon un point de vue connexionniste. Le contenu de la mémoire à un instant donné (mémoire à court terme) correspond à l'ensemble des éléments qui ont atteint leur seuil d'activation. Les processus d'activation (et d'inhibition) des éléments du réseau peuvent être *automatiques* ou *contrôlés*, ce qui rappelle l'importance de cette opposition.

Selon Barrouillet (1996), les modèles qui renvoient à la première de ces métaphores sont le modèle de Case (1992), le modèle de Daneman et Carpenter (1983) et le modèle de Baddeley (2003). Les modèles mentionnés par Barrouillet comme relevant de la métaphore énergétique sont le modèle de Just et Carpenter (1992), le modèle de Engle (1992), le modèle ACT-R (Anderson et al., 2004) et le modèle de la mémoire de travail à long terme (Ericsson & Kintsch, 1995). La différence principale entre ces approches tient au fait que dans le premier cas la mémoire de travail est considérée comme une structure spécifique,

alors que dans le second cas c'est le fonctionnement des processus attentionnel qui fait émerger une mémoire de travail. Dans le second cas, certains auteurs en viennent à abandonner la notion de ressources (Ericsson & Kintsch, 1995).

Les sections suivantes permettent de présenter plus en détail ces conceptions de la charge cognitive, d'abord selon les points de vue classiques des travaux fondateurs, puis sous l'angle des modèles de ressources, ensuite selon l'approche plus récente du développement des modèles computationnels, puis enfin dans une perspective de gestion de l'activité.

A Les prémisses d'une notion de capacité limitée du système cognitif

- ***La période réfractaire psychologique***

Les auteurs font généralement remonter la notion de charge mentale aux premières études sur la *période réfractaire psychologique* (*e.g.* Leplat, 2001; Richard, 1996) qui ont donné lieu à l'hypothèse d'un canal unique de traitement de l'information (Welford, 1952). Selon cette hypothèse, les signaux sont traités séquentiellement par l'individu, dans l'ordre d'arrivée, et la réponse à un second signal est différée si la réponse au premier signal est en cours. La notion est empruntée à la physiologie et à la *« période réfractaire »* associée à l'excitabilité des axones neuronaux (*Cf.* Chanquoy, Tricot & Sweller, 2007, p. 13).

L'hypothèse d'une *période réfractaire psychologique* repose sur la décomposition élémentaire d'une tâche de traitement d'information. Celle-ci est réduite au traitement de deux informations (ou signaux) successives ; et c'est l'interférence d'une information sur la transmission de l'autre qui intéresse les théoriciens. Cette hypothèse fournit un canevas pour l'analyse des situations de double tâche.

- ***La théorie de l'information***

Dans sa présentation d'une histoire de la charge mentale, Leplat (2002) associe cette hypothèse de la *période réfractaire psychologique* au modèle de la *théorie de l'information* (Ombredane & Faverge, 1955) comme bases historiques de la notion de charge mentale. D'après la *théorie de l'information*, il est possible d'assimiler l'humain, par hypothèse, à un canal de transmission d'information.

Leplat (2002) commente ainsi la Figure 4-1 :

> *« Soit H, la quantité d'information émise, R, la quantité d'information transmise. Jusqu'à une certaine quantité H_0, R = H ; si H > H_0, alors R < H. Dans ce dernier cas R/H < 1 : on pourrait parler ici de surcharge opérationnelle. R = H_0 a pu être interprété comme la capacité limite du canal. »*

Le graphique de la Figure 4-1 représente l'augmentation linéaire de la quantité d'information transmise jusqu'à H_0, puis la stagnation de cette valeur à partir de H_0. Il ne s'agit pas, cette fois, d'une décomposition de la tâche – réduite à sa forme élémentaire – mais d'une description quantitative des informations émises/transmises au cours d'une tâche continue dédiée à cette transmission. Cette description suppose l'équivalence des *informations* entre elles puisque celles-ci sont toutes regroupées dans une catégorie unique. L'humain est

assimilé à un canal qui reçoit ces informations de façon indifférenciée. Cette hypothèse a donc des limites évidentes, mais elle permet de poser des hypothèses expérimentales claires et l'utilisation d'une échelle quantitative permet de définir des seuils.

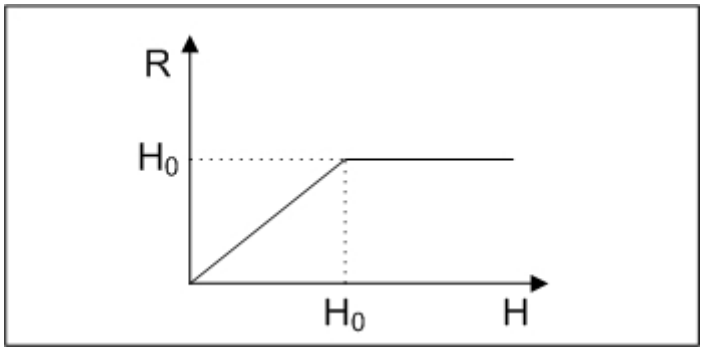


Figure 4-1 : Représentation graphique de la théorie de l'information (fonction : R = f(H))

- ***La double tâche et la mesure de l'attention***

Le paradigme expérimental de la double tâche consiste à faire exécuter deux tâches à un individu unique et au cours d'une période de temps limitée. Ce paradigme remonte à l'expérience *princeps* de Posner et Boies (1971). Camus (1998) présente ainsi cette expérience :

> *« Le sujet doit juger de l'identité de deux lettres présentées successivement (réponse, sans erreur et la plus rapide possible). Il doit aussi appuyer, le plus vite possible, sur un bouton en réponse à un signal sonore (on mesure le temps de réaction, TR, à cette tâche secondaire). Ce signal intervient à des moments différents lors de la présentation de la tâche principale. Par rapport à une mesure de référence où chaque tâche est exécutée séparément, l'allongement des temps de réaction à la tâche secondaire (audio motrice) constitue un indicateur de l'accroissement des ressources consommées par la tâche principale. Plus la tâche principale consomme de ressources et moins il en reste pour la tâche secondaire, en conséquence le TR moteur est proportionnellement allongé. » (page 160)*

Les deux tâches font appel à l'attention soutenue du sujet pendant toute la durée de l'expérience. L'interprétation de ce paradigme est basée sur l'idée que la performance à la tâche secondaire est une fonction de la difficulté de la tâche principale. Les deux tâches font appel aux ressources de l'individu. Les ressources doivent être divisées entre ces deux traitements. Le postulat sous-jacent est que la quantité de ressources attribuées à chacune des tâches est un prédicteur des performances obtenues.

Ce paradigme est très utilisé. Il a fournit une base expérimentale pour le développement et la mise à l'épreuve des modèles de l'attention qui sont présentés dans les pages qui suivent.

- ***La mémoire de travail***

Enfin, les travaux sur la *mémoire à court terme* d'abord (notamment Broadbent, 1958; Miller, 1956), puis sur la *mémoire de travail* (notamment Baddeley & Hitch, 1974) ont eu un impact important sur la conception du système cognitif comme un canal de traitement de l'information à capacité limitée.

L'article *princeps* dans ce domaine est l'article de Miller (1956) : *« Le nombre magique sept »*, qui stipule que nos capacités de mémoires sont limitées à un nombre fixé d'éléments (ou *« chunks »*), d'environ sept (plus ou moins deux) dans le cas de la mémorisation de listes de mots. Dans cet article, les arguments de l'auteur consistent à assimiler tout type d'unité de mesure dans une catégorie unique : *« information »,* qui peut alors être considérée comme une quantité sans dimension (*« a dimensionless quantity »*, Miller, 1956, p. 81). Cet argument est proche de celui qui est supposé par l'emploi de la théorie de l'information (*Cf.* page précédente). L'auteur base la notion d'*empan de mémoire immédiate* sur cette assimilation et il base la notion de *« chunk »* sur cet empan. Un *chunk* correspond à un emplacement de la mémoire à court terme (selon la métaphore spatio-temporelle) par opposition à un *« bit »* qui correspond à l'unité discriminée par le système perceptif. L'auteur stipule que :

> *« Je peux dire que le nombre de bits d'information est constant pour le jugement absolu et que le nombre de chunks d'information est constant pour la mémoire immédiate. » (Miller, 1956, page 92. Traduction libre).*

La différence entre ces deux notions suppose une opération de recodage au cours de laquelle les bits d'information sont assimilés. C'est la preuve faite de la nécessité de ces opérations de recodage qui a donné lieu au retentissement de cet article puisque ces opérations correspondent à des processus cognitifs à étudier.

Le cadre expérimental fourni ici permet l'analyse d'une tâche unique de mémorisation des informations. Comme cela a été noté au chapitre 3 (en conclusion de la présentation des théories de l'apprentissage multimédia) ce paradigme ne fournit pas l'analyse complète des activités nécessaires à l'accomplissement d'une tâche d'apprentissage. Sur le principe de la théorie de l'information, il s'agit d'assimiler, par hypothèse, l'humain à un canal de traitement des informations.

- ***Conclusion***

Ces éléments fournissent les bases de l'analyse de la charge mentale/cognitive des individus. Les modèles de ressources ont été proposés en s'appuyant sur ces notions.

B Les modèles de ressources

Les modèles de ressources correspondent à des tentatives de synthèse du fonctionnement du système cognitif selon une approche générale et unifiée. Ils utilisent les bases qui viennent d'être présentées, à partir desquelles les auteurs ont cherché des principes descriptifs utilisables pour tout type de tâche.

- ***Théorie d'une réserve de ressources unique***

Les conceptions modernes sur la notion de ressources prennent naissance avec l'ouvrage de Daniel Kahneman (1973) sur *l'effort et l'attention*. A la suite de cet ouvrage, trois articles importants ont participé au développement du modèle : Norman et Bobrow (1975), Navon et Gopher (1979) et Wickens (1984).

La première théorie, celle de Kahneman (1973), propose une réserve de ressources unique. La performance à une tâche dépend alors des ressources qui lui sont allouées dans cette réserve. La réserve de ressources est limitée et dépend du niveau d'éveil. Une stratégie d'allocation détermine la capacité allouée à une tâche. Elle dépend des caractéristiques individuelles et des influences motivationnelles.

Norman et Bobrow (1975) introduisent une distinction supplémentaire en proposant que les processus peuvent être *limités par les ressources* (*« resource-limited »*) ou *par les données* (*« data-limited »*). En effet, la performance ne dépend pas nécessairement des ressources investies par l'individu, mais peut également dépendre de la qualité des données reçues.

Cette notion a ensuite été étendue par Navon et Gopher (1979) qui proposent l'expression *'paramètres sujet-tâche'* (*« subject-task parameters »*) pour désigner à la fois les caractéristiques de la tâche et les propriétés permanentes et transitoires du sujet. Les auteurs n'incluent plus seulement les processus *limités par les données*, mais également les *interactions* qui peuvent apparaître *entre la tâche et l'opérateur*, notamment en fonction de son niveau d'*expertise*. Leur modèle permet de prendre en compte les conditions dans lesquelles les ressources sont mobilisées et allouées à la tâche. Ils y incluent également les fluctuations de ressources disponibles, liées à l'éveil, et définissent la *'charge mentale'* comme une *réponse émotionnelle aux exigences d'une tâche*. Ils désignent par là le niveau d'investissement du sujet et ses fluctuations (motivation, ajustement de l'effort).

- ***Le modèle de ressources multiples***

Navon et Gopher (1979) argumentaient déjà en faveur d'un processeur à multiples canaux, chacun ayant sa propre capacité, mais celle-ci pouvant être partagée. La modification de la difficulté d'une tâche peut entraîner des modifications structurelles sur le processus de traitement. Ainsi, la performance n'est pas une fonction linéaire de la difficulté ; et il est donc nécessaire de s'intéresser aux effets locaux et globaux entraînés par la modification d'une tâche. Pour répondre à ce besoin, Wickens (1984) propose un modèle basé sur trois axes qui permettent de différencier les principaux *pools de ressources*. Ces trois axes sont :

(1) Le *code de traitement* des informations, qui peut être *'verbal'* ou *'spatial'* ;

(2) La *modalité de traitement* des informations, *'visuelle'* ou *'auditive'* dans le modèle original, mais les autres modalités peuvent également y être intégrées ;

(3) Le *niveau de traitement* (ou stade) des informations, qui inclut le *'traitement perceptif'*, le *'traitement central'* et le *'traitement de la réponse*'.

La projection de ces trois axes sur une figure forme un cube dont les parties représentent les pools de ressources cognitives humaines. Cette représentation est nommée *modèle de ressources multiples* (Figure 4-2). Dans ce modèle, les boucles perception-action peuvent être représentées sous formes de flèches prenant des directions différentes à l'intérieur du cube. Par exemple, le langage oral est perçu par la modalité auditive et encodé selon le code verbal (en bas à droite de la face « encodage »), la réponse est vocale et également encodée

selon le code verbal (à droite de la face « réponse »). Des tâches visuelles avec réponse manuelle gravitent dans la face gauche du cube. Wickens (2002) précise que les tâches qui supposent des changement de code impliquent des traitements centraux plus coûteux.

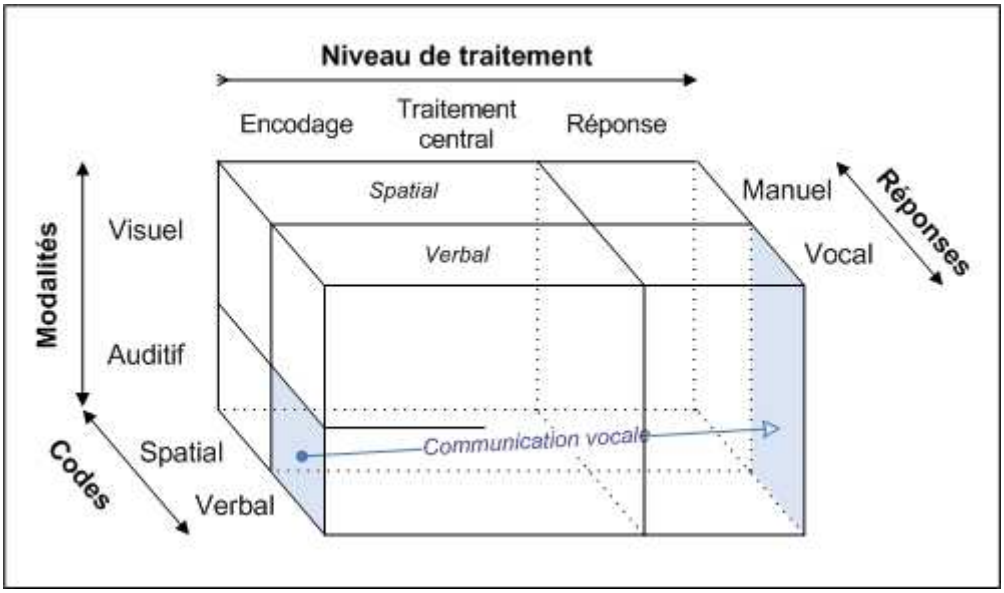


Figure 4-2 : Le modèle de ressources multiples (Wickens, 1984)

Wickens (1984) reconnaît lui-même certaines imprécisions de son modèle eu égard à celui de Navon et Gopher (1979). Cependant, le *modèle de ressources multiples* permet de prédire certains effets liés aux modes de perception et de réponse qui ne peuvent être expliqués dans une conception d'un pool de ressources unique. D'après Wickens (1987), en condition de double tâche un partage temporel parfait n'intervient que s'il n'y a aucun recouvrement entre les demandes de ressources liées aux deux tâches. Mais cette affirmation a été remise en cause depuis (voir Anderson et al., 2004) et des auteurs ont montré que le partage temporel entre plusieurs tâches dépend du niveau de pratique de ces tâches (*e.g.* Hazeltine, Teague & Ivry, 2002; Schumacher et al., 2001).

La conséquence de ce modèle est que les ressources ne sont plus définies simplement comme une réserve d'énergie simple, mais comme le couplage d'une structure de traitement d'information et de l'énergie disponible pour cette structure. C'est sous cette condition que l'on peut prétendre recourir à la *métaphore énergétique* de Barrouillet (1996).

- ***Critique des modèles de ressources***

Une critique importante des théories du domaine a été formulée par l'un de ses auteurs principaux (Navon, 1984) dans un article au titre évocateur (*« Ressources : une soupe de cailloux théorique »*). Cette critique consiste à remarquer que la notion de ressources ne permet pas de constituer un modèle du comportement. L'auteur concluait ainsi son article :

> *« La morale que j'espère faire émerger de ma discussion est que les tentatives de mesurer la charge mentale, d'identifier les réserves de ressources, de prédire les interférences entre tâches par des fonctions de performance de ressources, ou d'incorporer l'allocation des ressources dans un modèle des processus du comportement, peuvent s'avérer aussi décevantes que le serait la tentative d'isoler*

> *dans l'esprit humain les analogues des composants fonctionnels d'un ordinateur digital. »* (Navon, 1984, p. 232. Traduction libre.)

Ainsi, l'auteur indique que l'étude des ressources prend le parti d'une description structuraliste, mais que celle-ci est insuffisante. Une approche fonctionnelle s'impose.

Ce domaine montre cependant que la charge mentale occupe un rôle de régulateur (Navon & Gopher, 1979), qui relève justement d'une approche fonctionnelle. Des développements plus récents montrent qu'il est important de ne pas considérer la *charge mentale* isolément de la *représentation de la situation* (*« situation awareness »*) pour guider la conception des systèmes automatiques (Tsang & Vidulich, 2005; Tsang & Wilson, 1997). La notion de charge mentale peut être dépassée car elle n'est pas suffisante pour expliquer les diverses causalités impliquées dans le processus comportemental.

C Les modèles computationnels

Les modèles computationnels représentent un niveau d'évolution supplémentaire et ils permettent de réinterpréter les conceptions issues des modèles de ressources. Ils sont utilisés en psychologie pour la simulation des processus cognitifs individuels. Ils relèvent de techniques équivalentes à celles qui ont été évoquées au chapitre 2 (représentation des états mentaux, programmation logique). Ces modèles permettent de modéliser le fonctionnement cognitif général et d'intégrer les processus spécialisés selon des techniques symboliques et connexionnistes (sub-symboliques). Ils peuvent permettre d'interpréter des processus spécialisés tels que la mémoire (voir par exemple, Kieras, Meyer, Mueller, & Seymour, 2003; Lovett, Reder & Lebiere, 2003). Mais l'intérêt de ces approches est surtout de fournir un modèle de l'individu qui peut être utilisé, soit pour modéliser la réalisation de certaines tâches (pour prédire les performances), soit pour concevoir des systèmes interactifs capables d'adopter un comportement adaptatif collaboratif.

Les auteurs s'appuient sur le modèle de ressources multiples (Wickens, 1984) et réfutent l'idée d'un goulot d'étranglement unique (Welford, 1952). Ils affirment au contraire que la *charge de travail mentale* est un *phénomène distribué* dans l'ensemble du système, qui ne se limite pas aux traitements cognitifs. Kieras et Meyer (1997, p. 396) expriment clairement cette position dans la présentation de leur modèle :

> *« Nous avons présumé que les limitations des capacités humaines sont toutes structurelles ; c'est-à-dire que la performance à des tâches peut être limitée par les contraintes sur les périphériques perceptifs et sur les mécanismes moteurs plutôt que sur les envahissantes limites des traitements cognitifs. Ainsi, la stratégie exécutive a la responsabilité d'aller à la rencontre de la performance face aux exigences des tâches malgré ces limitations structurelles. »* (Traduction libre)

Selon ce point de vue, la notion de limite du système cognitif doit être appliquée individuellement à chacun des processus et mécanismes impliqués par la production des réponses de l'individu. Trois des modèles conçus pour formaliser ces mécanismes sont évoqués ici (3CAPS, ACT-R et EPIC) car ils permettent de relever certaines des propositions

importantes pour la compréhension de la notion de charge cognitive et pour identifier le type de modélisation qui est privilégié dans une perspective de conception.

- ***3CAPS***

Le modèle 3CAPS (Just & Carpenter, 1992) est dédié au problème des ressources. Il a été proposé pour expliquer les différences individuelles de la capacité de la mémoire de travail dans le domaine de la compréhension du langage. Il est utilisé pour mesurer et contrôler la charge mentale au cours de l'exécution de différentes tâches.

Les résultats principaux (*Cf.* Just, Carpenter & Miyake, 2003) indiquent que lors d'une double tâche (*e.g.* conduite automobile et vérification de phrases présentées auditivement) les ressources consommées pour chaque tâche sont moins importantes (de l'ordre de 25 à 40 %) que quand chacune est réalisée individuellement. Inversement, certaines zones cérébrales qui ne sont activées par l'exécution d'aucune des deux tâches exécutées individuellement sont activées lors de l'exécution simultanée. Pour les auteurs, ces résultats montrent que des interactions entre les tâches interviennent et que la consommation de ressources n'est pas simplement additive.

Ce modèle a été utilisé pour mesurer et contrôler la charge cognitive en contexte applicatif. Just, Carpenter et Miyake (2003) mentionnent deux de ces applications. L'une concerne l'interaction téléphonique (Huguenard, Lerch, Junker, Patz, & Kass, 1997), l'autre porte sur une tâche de contrôle du trafic aérien (Byrne & Bovair, 1997). La première de ces applications est présentée en détail à la fin de ce chapitre.

- ***ACT-R et EPIC***

ACT-R (*"Adaptive Control of Thought – Rationale"*, Anderson et al., 2004) et EPIC (*"Executive Process – Interactive Control"*, Kieras & Meyer, 1997) sont les deux modèles computationnels les plus cités en psychologie cognitive. Ces deux modèles offrent des perspectives complémentaires pour la description du fonctionnement cognitif. ACT-R focalise sur le fonctionnement de haut niveau (traitement central) et privilégie moins les aspects perceptifs et moteurs pour lesquels EPIC est plus complet. Les auteurs font de nombreuses références mutuelles. Cependant, les deux modèles ont des origines différentes. La théorie ACT a été développée spécifiquement dès les premiers travaux d'Anderson (Anderson & Bower, 1973) et se base sur la distinction entre connaissances déclaratives et procédurales (Anderson, 1982). EPIC a été proposé plus récemment et s'appuie notamment sur le modèle de Wickens (1984). Les schémas fonctionnels correspondant à ces deux modèles sont présentés dans la Figure 4-3 (page 102) :

- Le ***modèle ACT-R 5.0*** (Figure 4-3, à droite) permet de représenter l'architecture des traitements cognitifs nécessaires à l'exécution d'une tâche et de faire l'appariement (*« mapping »*) entre ces traitements et les aires cérébrales qui en sont responsables (Anderson, Qin, Sohn, Stenger, & Carter, 2003). Chaque module représenté dans la figure présente d'abord le nom du traitement, puis l'aire cérébrale entre parenthèses. La théorie permet de décrire le fonctionnement du système. Au centre de l'architecture, le

système de production est responsable de la sélection des informations émanant des autres modules, de la résolution des conflits et du contrôle de l'exécution des actions. Ce module a un fonctionnement cyclique, responsable de la coordination de l'ensemble des éléments de l'architecture. Un cycle a une durée de 50 millisecondes. A chaque cycle, le niveau d'activation des connaissances émanant des différents modules est mis à jour, ainsi que les buts (plans d'action) en fonction du niveau d'avancement de la tâche (voir Anderson et al., 2004). Pour le fonctionnement des systèmes perceptifs et moteurs, ces auteurs ont adopté le fonctionnement proposé par Meyer et Kieras (1997) dans le cadre d'EPIC.

La théorie ACT repose sur deux systèmes de mémoire distincts (Anderson, 1982) : (1) la *mémoire déclarative* encode les connaissances sur les faits et événements qui sont organisés en structures appelées *« chunks »* ; (2) la *mémoire procédurale* est organisée sous forme de *« règles de production »* qui réfèrent au *but* courant et aux *chunks* en mémoire à long terme. Chaque règle de production est définie par un *coût* (temps nécessaire à la réalisation de la procédure) et une *probabilité de succès*.

Du point de vue méthodologique, lors de la conception d'une application particulière, la structure de buts doit être décrite à l'avance par les concepteurs de l'application. Selon Anderson (1996) cette description est faite préalablement lors d'une analyse des tâches. Mais cette méthode n'est pas détaillée par l'auteur.

- Le ***modèle EPIC*** (Figure 4-3, à gauche) est plus directement impliqué dans la prédiction de la performance en fonction à la fois des processus perceptifs et moteurs et des processus cognitifs. Meyer et Kieras (1997) s'appuient sur une série d'arguments favorables au développement des modèles computationnels : (1) Ils permettent de diversifier les processus étudiés. (2) Ils offrent un cadre de travail structurellement stable, sur des bases techniques. (3) Ils permettent la mise à l'épreuve du modèle conçu dans le cadre de l'interaction avec un système (IHM). (4) Ils permettent d'incorporer les processus perceptifs et moteurs en plus des processus cognitifs. (5) Ces modèles permettent l'analyse des processus exécutifs (stratégies de réalisation des tâches) à l'aide de formalismes tels que GOMS (Card, Moran & Newell, 1983) ou la méthode d'analyse de chemin critique (*"critical path analysis"*, *e.g.* Baber & Mellor, 2001). (6) Enfin, ces modèles permettent d'omettre l'hypothèse de capacité limitée. Les auteurs considèrent qu'elle n'est pas nécessaire. Sur ce point, ils s'appuient sur une série de remarques d'Allport (1980, p. 117-118) :

  « Evidemment une difficulté est de savoir quand faire appel à une compétition pour un pool de ressources unique. (…) Une fois que l'on accepte l'idée générale d'une capacité de traitement comme hypothèse de travail, il devient facilement tentant de considérer, sans autre procès, que toute interférence en contexte de double tâche est le résultat d'une compétition pour les ressources, pour ''l'attention''. (…) La théorie, au moins dans son application, apparaît être totalement circulaire. (…) Le résultat est une stratégie de recherche qui ne peut rien faire d'autre que se mordre la queue. (…) Cela a été une heuristique singulièrement improductive pour la découverte des contraintes

architecturales qui agissent sur les processus psychologiques concurrents. (…) Cela bloque la curiosité en donnant l'apparence de fournir une explication avant même d'obtenir des données. » (Traduction libre)

Kieras et Meyer (1997) proposent d'intégrer tous les processus ayant un rôle dans la prédiction de la performance. Ils considèrent que : *« le système moteur humain est plutôt complexe dans son fonctionnement et interagit fortement avec les systèmes cognitif et perceptif. »*. Leur modèle (à gauche) est ainsi un élargissement du modèle proposé dans le cadre d'ACT-R (à droite). Les processus externes, liés à l'exécution de la tâche sont mieux pris en compte : les contraintes fonctionnelles liées aux organes sensoriels et moteurs permettent d'expliquer la performance à la tâche (*e.g.* tâche de recherche d'un item dans une liste déroulante). Ce modèle permet des analyses très fines du comportement.

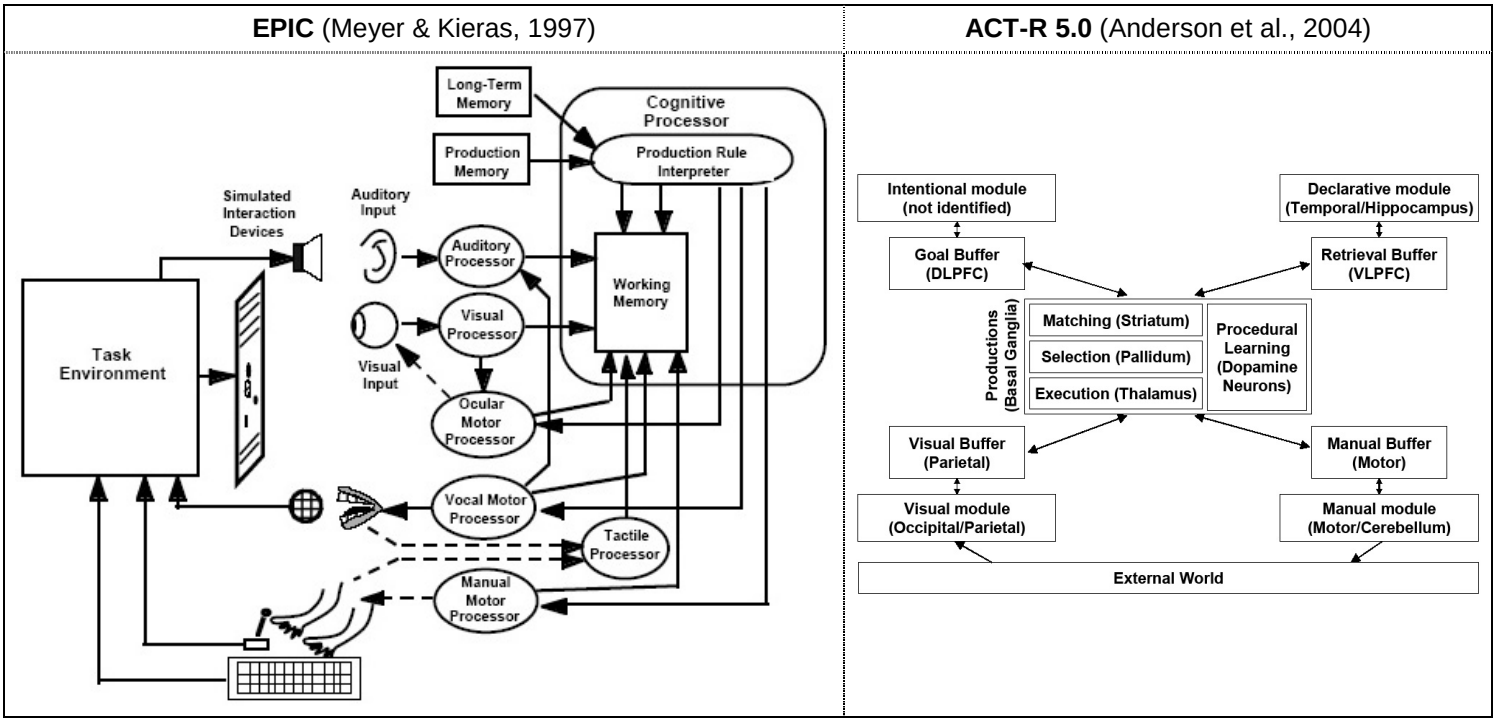


Figure 4-3 : ACT-R et EPIC

- ***Conclusion***

Pour ces auteurs, les résultats qu'ils obtiennent contredisent les modèles de ressources. Ils montrent comment la pratique permet de convertir les informations déclaratives sur une tâche (*i.e.* la formulation explicites des règles de production) en connaissances procédurales (*i.e.* les règles de production automatisées) (Anderson, 1982). Après cette conversion, les tâches peuvent être exécutées en parallèle. De plus, les interactions entre tâches n'entrainaient pas des consommations de ressources additives (Just, Carpenter & Miyake, 2003). D'après Kieras et Meyer (2000), lorsque l'arrangement temporel des stimuli et les consignes de priorité des tâches sont manipulés, les résultats montrent que la stratégie de contrôle exécutif (ordonnancement des tâches) est susceptible d'induire une interférence. De plus, des différences individuelles systématiques apparaissent, dues à la mise en place de stratégies spécifiques, rapidement apprises par les sujets, d'ordonnancement temporel des tâches.

Ainsi, dans ces modèles, c'est l'organisation parallèle des processus qui est utilisée comme base explicative de la performance plutôt qu'une capacité générale. Pour Just, Carpenter et Miyake (2003) la charge ne peut être définie clairement car elle pose d'incessants problèmes de granularité lors de l'observation entre approche généraliste du système nerveux central et focalisation sur les centres nerveux qui participent activement aux différents traitements.

La série d'arguments de Meyer et Kieras (1997) en faveur du développement des modèles computationnels est difficile à ignorer car ces modèles offrent des solutions à de nombreuses limites de l'approche traditionnelle de la psychologie expérimentale. Des processus divers peuvent être pris en compte, formalisés et testés. La complémentarité des modèles montre qu'il est nécessaire d'intégrer plusieurs niveaux d'analyse pour expliquer l'activité de l'opérateur. Schématiquement, on distingue un niveau cognitif (ACT-R) et un niveau perceptivo-moteur (EPIC). De tels modèles permettent la mise en place d'une méthode de travail plus analytique, plus systématique et, par là, plus complète. Finalement, les auteurs qui travaillent au développement de ces modèles tendent à abandonner progressivement la notion de charge mentale ou de charge cognitive pour privilégier cette approche analytique.

D Gestion de l'activité

Les éléments présentés jusqu'ici discutent principalement la charge mentale sur la base des prémisses qui ont été proposées au début de cette partie et sous l'angle du fonctionnement de l'individu (*'théories de la boîte noire'*, Vidulich, 2003), *i.e.* sous l'angle de sa *charge cognitive*. Des points de vue orientés sur l'analyse des tâches et sur la gestion de l'activité de l'opérateur ont également été développés, notamment en ergonomie (*e.g.* Spérandio, 1972, 1977) et dans le cadre du développement des *« human factors »* américains (voir, Andre, 2001; Parasuraman & Hancock, 2001; Wickens, 2001). Dans ce cas, les auteurs parlent plus volontiers de *'charge de travail'* (*« workload »*) ou de *'charge de travail mentale'* (*« mental workload »*).

L'ergonomie francophone a donné lieu à de nombreux travaux orientés sur l'analyse de l'activité, où la notion générale de charge mentale a pu être analysée sur des bases objectives liées au déroulement de la tâche (voir, Leplat, 2002; Theureau, 2002). Ces travaux tendent à conduire au rejet d'une notion trop générale de charge qui ne permet de traiter que de certains cas limites (*i.e.* identification des conditions limites d'une tâche, définition de seuils de difficulté). De Montmollin (1997, article "charge de travail") illustre par *« la parabole du cancre »* le fait que tous les individus n'ont pas le même rapport à l'effort et que celui-ci ne prédit pas la performance. Les différents individus *« peinent plus ou moins efficacement »*. Les travaux s'orientent sur l'identification des cas limites, c'est-à-dire des *surcharges de travail*. Sur ce point, une littérature s'est développée autour de la suractivité des cadres et de l'omniprésence des moyens de communication, sur le thème du *« syndrome de débordement cognitif »* (voir, par exemple, Lahlou, 2002) ou COS (*« Cognitive Overload Syndrome »*). Ce courant tend par ailleurs à prendre en compte l'approche de la cognition située (Suchman,

1987) qui ne privilégie pas des notions générales telles que celle de charge de travail (voir, par exemple, Le Guilcher & Villame, 2002).

Les *« human factors »* ont accompagné le développement des approches théoriques déjà évoquées (modèles de ressources, etc.) et ont participé au développement d'outils de mesure et d'analyse basés sur des approches diverses (en ingénierie, physiologie, psychologie), surtout dans le domaine du trafic aérien. Les auteurs se basent sur des synthèses théoriques dont l'objectif est d'intégrer les différents facteurs susceptibles d'influencer la charge de travail. La Figure 4-4 présente le schéma de synthèse proposé par Hart et Staveland (1988).

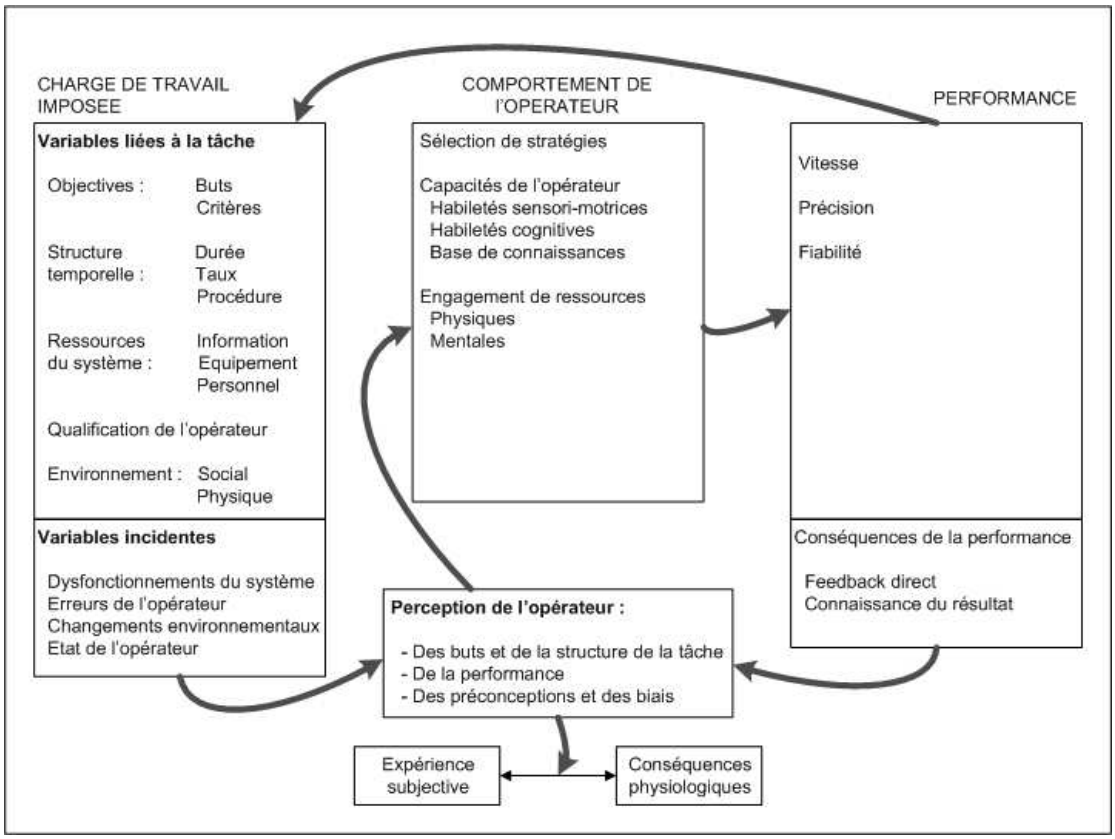


Figure 4-4 : Cadre conceptuel - Relations entre les variables qui influencent la charge

Pour Hart et Staveland (1988), la charge de travail est un « *concept hypothétique qui représente le coût engagé par l'opérateur humain pour atteindre un niveau de performance particulier »* (p. 140). Ces auteurs considèrent que leur définition est centrée-humain (*« human-centered »*) mais ils proposent un cadre conceptuel qui intègre les facteurs internes et externes. A gauche de la figure, la *'charge de travail imposée'* réfère à la situation rencontrée par l'opérateur. La tâche y est décrite en fonction de critères objectifs. Ces aspects sont nommés *« contrainte »* par Spérandio (1980), par opposition à *« l'astreinte »*[1] qui en résulte pour l'opérateur et qui correspond, dans la Figure 4-4, au *'comportement de l'opérateur'*, au centre. D'après ce schéma, les contraintes imposées par la tâche sont perçues par l'opérateur de sorte qu'il s'astreint à un certain comportement en fonction de ses capacités et de ses engagements et qu'il parvient à obtenir une certaine performance (à

[1] Pour Spérandio (1980) c'est l'astreinte qui définit la charge de travail.

droite). Celle-ci est directement perçue par l'opérateur (boucle 1, en bas) et elle modifie progressivement les contraintes (boucle 2, en haut). Spérandio (1972) insistait également sur la distinction de ces deux boucles de rétroaction. On voit dans ce schéma que c'est dans le déroulement cyclique de l'activité que les différents facteurs impliqués influencent la charge.

De même, Wickens (2001) indique que la charge doit être *triangulée* entre : (1) difficulté de la tâche, (2) compétence de l'opérateur (*« operator skill »*) et (3) performance. Cet auteur fait un parallèle avec la notion de *'représentation de la situation'*[1] (*« situation awareness »*, voir aussi Tsang & Vidulich, 2005; Vidulich, 2003; Wickens et al., 2005). Cette notion renvoie à la compréhension que l'opérateur a du problème en court de traitement. Elle renvoie également à un grand nombre de variables en jeu dans la gestion de l'activité, mais sa nature qualitative la distingue de la notion de charge de travail. Pour Wickens (2001), l'émergence de ces concepts est liée aux préoccupations liées à la sécurité en ingénierie et au besoin de définir des *'lignes rouges'* basées sur des valeurs quantifiables. Cependant, les notions de *charge de travail* et de *représentation de la situation* permettent d'aborder des aspects quantitatifs et qualitatifs qui sont complémentaires. L'intérêt de ces notions est de donner lieu au développement de modèles prédictifs, de nature généraliste, dédiés à la prédiction de la performance humaine. De même, pour un auteur comme Spérandio (1972), la finalité de l'étude de la charge de travail réside surtout dans l'étude de la *variation des processus opératoires*, c'est-à-dire dans l'étude de l'activité déployée par l'opérateur en fonction de ses compétences dans l'exécution d'une tâche finalisée.

E Conclusion

En termes de définition, les auteurs déplorent le flou de la définition de la notion de charge mentale. Mais ces problèmes n'émanent pas d'un manque de volonté. La notion s'avère plutôt sur-définie que sous-définie. Le problème qui se pose réellement en ce qui concerne cette notion est clairement expliqué par Theureau (2002) pour qui elle agit dans le champ scientifique comme un prétexte théorique qui permet d'étudier des problèmes concrets. Dans ce cas, le problème ne serait pas tant de bien la définir que de comprendre qu'elle ne fournit qu'une première approche du phénomène étudié, qui ne saurait en révéler la clé et qui doit être dépassée.

4.1.3 Problèmes de mesure

Comme l'ont remarqué Tsang et Wilson (2004), les principaux progrès en ce qui concerne la charge mentale ont d'abord porté sur son évaluation. Cette partie permet d'évoquer rapidement la diversité des indicateurs utilisés, pour présenter ensuite les deux questionnaires d'évaluation subjective les plus représentatifs.

[1] Littéralement, *« situation awareness »* se traduit *« conscience de la situation »*. Mais la traduction par *'représentation de la situation'* est utilisée ici pour renvoyer à une notion plus classique.

A Diversité des indicateurs

Une mesure directe et complète de la charge mentale d'un individu au cours de la réalisation d'une tâche supposerait l'enregistrement de l'ensemble de son activité cérébrale au cours de cette tâche (Just, Carpenter et Miyake, 2003). Mais cet objectif n'est généralement pas atteignable et même l'emploi d'un dispositif d'imagerie cérébrale suppose de se concentrer sur un aspect de cette activité. Les indicateurs de charge généralement utilisés sont indirects car ils s'appuient sur des valeurs qui tendent à varier avec l'activité cérébrale et qui permettent d'en donner une estimation. De nombreux indicateurs ont été proposés. Ils sont généralement classés en trois catégories : (1) les indicateurs physiologiques, (2) les mesures de performance et (3) les questionnaires d'évaluation subjective (*Cf.* Tableau 4-1).

Tableau 4-1 : Diversité des indicateurs de charge

Type de mesures	Indicateurs	
Indicateurs physiologiques	*- Battements cardiaques* *- Calorimétrie* *- Réponse électrodermale*	*- Diamètre pupillaire* *- EEG (P300, etc.)*
Mesures de performance	*- Efficacité* *- Productivité* *- Taux de réussite / échec*	*- Taux d'erreur* *- Temps d'exécution*
Questionnaires subjectifs	*- Echelle de Bedford* *- Echelle Psychophysique* *- Echelle de Cooper-Harper* *- SWAT*	*- NASA-TLX* *- Workload Profile* *- SMWS*

Ces indicateurs ont tous tendance à varier en fonction de la difficulté de la tâche. Mais ils ne sont pas tous adaptés aux mêmes tâches. Par exemple, les indicateurs physiologiques sont très sensibles à un travail manuel. Si un travail fait intervenir une charge physique importante, ils ne peuvent être utilisés comme indicateurs de l'activité mentale. De même, les mesures de performance peuvent varier en fonction du niveau d'expertise ou d'éveil de l'opérateur, de sorte que ces mesures supposent des contrôles expérimentaux (population testée, calibration et scénarisation des tâches) qui en relativisent la portée. Les questionnaires subjectifs ont le même défaut. Ils ne permettent pas de faire un diagnostic sur une tâche (*la tâche est-elle trop difficile ou non ?*), mais ont une valeur comparative (*la tâche est-elle plus difficile dans une condition que dans une autre ?*). La mesure du diamètre pupillaire est une mesure fidèle (*e.g.* Iqbal, Zheng & Bailey, 2004), mais elle suppose l'emploi d'un dispositif d'*eye tracking*.

Etant donné la facilité d'usage, les techniques les plus fréquemment utilisées sont les *mesures de performance* et les *questionnaires d'évaluation subjective*. Les premières sont spécifiques à la tâche évaluée et à la méthodologie utilisée (*e.g.* paradigme en double tâche). Dans le cas du dialogue, elles renvoient aux descripteurs du processus collaboratif, qui ont été présentés au chapitre 1. Les secondes sont basées sur des questionnaires prédéfinis, qui utilisent différents types de catégorisation des contraintes associées à la tâche ou des ressources de l'individu. Elles sont abordées dans la section suivante.

B Deux questionnaires de mesure subjective : NASA-TLX et 'Workload Profile'

Les premiers questionnaires subjectifs ont été développés dans le cadre de l'étude du pilotage et du contrôle aérien, avec notamment l'*échelle de 'Cooper-Harper'* (1969). D'autres solutions ont été proposées au cours du temps, spécialement dans ce domaine du trafic aérien, et ont donné lieu à une diversité de questionnaire : l'*échelle psychophysique*' (Gopher & Braune, 1984) *'l'échelle de Bedford'* (Roscoe, 1987), les questionnaires *'SWAT'* (Reid & Nygren, 1988), *'NASA-TLX'* (Hart & Staveland, 1988) et *'Workload Profile'* (Tsang & Velasquez, 1996). Certaines propositions ont également été faites dans le cadre des théories de l'apprentissage multimédia. Yeung, Lee, Pena et Ryde (2000) ont proposé un questionnaire supplémentaire (*'SMWS'*) ; et d'autres auteurs se basent sur des échelles de Likert simples (Van Merriënboer, Schuurman, De Croock, & Paas, 2002). L'avantage de ce type de mesure est de se baser sur la relation subjective de l'individu à la tâche qu'il réalise, puisque c'est cette relation qui définit la charge mentale. Mais il est également possible de considérer que cette subjectivité est source d'imprécision.

Seuls *'NASA-TLX'* et *'Workload Profile'* sont présentés dans cette partie car ils donnent lieu à des points de vue complémentaires et parce que leur utilisation est répandue.

- ***'NASA-TLX'***

'NASA-TLX' est le résultat d'une recherche de plusieurs années spécifiquement dédiée à son développement (Hart & Staveland, 1988). Ce questionnaire est le plus utilisé des outils d'évaluation subjective de la charge de travail. Certains auteurs en utilisent des versions simplifiées (voir, par exemple, Le Bigot, 2004).

Le modèle sous-jacent à TLX a été présenté dans la section précédente (Figure 4-4). La démarche de conception du questionnaire a consisté en plusieurs étapes qui ont permis de proposer un ensemble de qualificatifs de la charge, puis de sélectionner et d'associer les différentes notions entre elles. Une liste de 19 facteurs a été proposée à partir des éléments de la Figure 4-4. La première étape a consisté à demander à des personnes occupant différentes fonctions professionnelles d'évaluer l'équivalence de ces facteurs avec la charge de travail. Parmi les 19 facteurs, 14 ont été jugés équivalents à la charge par plus de 60% des participants. La seconde étape a consisté à demander à plusieurs groupes d'évaluer leur effort avec ces 14 facteurs dans une série de tâches de laboratoire et en simulateur de vol. Ce test a permis de faire des rapprochements entre les variables qui co-variaient fortement (*e.g.* difficulté de la tâche et complexité, effort et stress). A partir de ces éléments, dix échelles bipolaires ont été développées : (1) charge globale, (2) difficulté de la tâche, (3) pression temporelle, (4) performance propre, (5) effort physique, (6) effort mental, (7) frustration, (8) stress, (9) fatigue, (10) type d'activité (niveau d'abstraction, *Cf.* Rasmussen, 1983). Ces dix échelles ont ensuite été utilisées dans une série de 25 études. Des comparaisons internes aux différentes études (analyses de variance et corrélation avec la performance) ont permis de retenir 16 études dans lesquelles les conditions expérimentales donnaient lieu à des évaluations subjectives différentes. Une série d'analyses transversales à

ces 16 études a été menée (relativement au type de tâche, aux participants et à la performance) pour déterminer les facteurs les plus importants. Ces analyses ont permis d'isoler six dimensions principales :

1. *'Demande mentale'* : A quel point l'activité mentale est-elle requise ?
2. *'Demande physique'* : A quel point l'activité physique est-elle requise ?
3. *'Demande temporelle'* : Quelle pression temporelle est ressentie ?
4. *'Performance'* : Quel succès est ressenti ?
5. *'Effort'* : Quel effort est nécessaire ?
6. *'Niveau de frustration'* : Quelle frustration est ressentie ?

Mais les auteurs notent que l'évaluation de ces six échelles est insuffisante pour obtenir un score global de charge de travail. D'une part, la définition de la charge est variable d'un individu à l'autre. D'autre part, le poids associé à chaque source de charge est variable d'une tâche à l'autre. Une procédure supplémentaire est donc nécessaire pour pondérer la valeur des différentes échelles. Il s'agit d'une *procédure de comparaison par paires* (*« pair-wised comparison »*). Cette procédure consiste à utiliser les six échelles proposées et de les combiner par paires pour former tous les couples possibles (15 au total). Pour chacun de ces couples, le participant doit déterminer celle des deux échelles qui lui semble la plus importante. Chacune des six échelles obtient ainsi un score qui est utilisé pour pondérer l'évaluation faite sur cette échelle. *L'indice TLX* est une moyenne des évaluations des six échelles pondérée (respectivement) par ces six scores.

D'après Hart et Staveland (1988), seul cet indice général doit être utilisé comme qualificatif du concept théorique *charge de travail* (*« workload »*). Selon ce point de vue, les dimensions qui composent le questionnaire ne doivent pas être interprétées indépendamment les unes des autres car la charge de travail est vue comme une notion indépendante. Le questionnaire est multidimensionnel mais le concept théorique est unifié.

- ***'Workload Profile'***

'Workload Profile' a été proposé plus récemment (Tsang & Velasquez, 1996) dans le but de permettre l'évaluation subjective de l'aspect multidimensionnel de la charge de travail. La définition qui est donnée de la charge dans ce questionnaire la rapproche de la notion de *'charge cognitive'* (ou de *'charge de travail mentale'*) dans le sens où l'évaluation est focalisée sur « la boîte noire ». Les auteurs s'appuient sur le modèle de Wickens (1984) puisque celui-ci est basé sur un vaste ensemble de validations empiriques. Tsang et Velasquez (1996) supposent que l'utilisation sous-jacente de ce modèle peut permettre d'obtenir des évaluations dont la valeur serait diagnostique.

Seuls les pools de ressources sont utilisés pour l'évaluation. Ils définissent les capacités cognitives de l'individu et celui-ci doit évaluer la quantité utilisée (taux d'utilisation. Valeur entre 0 et 1) dans chacun de ces pools de ressources lors de l'accomplissement de la tâche.

Huit types de traitements cognitifs sont évalués (la Figure 4-5 les localise dans le *cube de Wickens*). Lors de la passation d'une expérience, les participants disposent d'une fiche explicative qui leur indique la signification de chacune des dimensions qui doivent être évaluées (*Cf.* Annexes). Cette fiche est nécessaire pour assurer une bonne compréhension des participants, pour leur permettre une bonne précision dans leurs évaluations.

1. Traitement perceptif / central ;
2. Traitement de la réponse ;
3. Traitement visuel ;
4. Traitement auditif ;
5. Traitement spatial ;
6. Traitement verbal ;
7. Réponse manuelle ;
8. Réponse vocale.

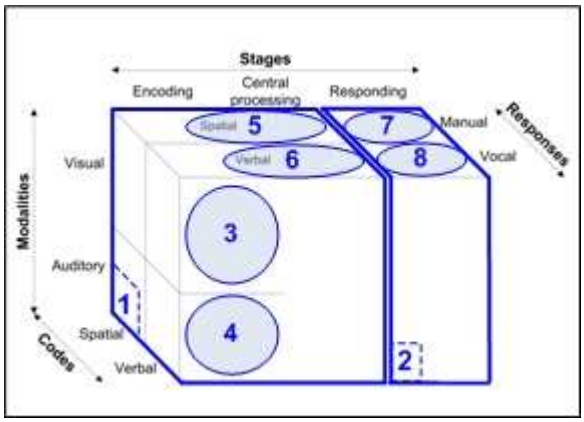


Figure 4-5 : Les traitements évalués avec WP

Ces évaluations permettent d'obtenir un *profile de charge de travail* (*« workload profile »*) relatif à la tâche, mais défini indépendamment des caractéristiques de la tâche puisqu'il ne repose que sur les ressources de l'individu. Tsang et Velasquez (1996) ont comparé ce questionnaire à deux méthodes d'évaluation unidimensionnelles (l'*échelle de Bedford'* et l'*échelle psychophysique'*) dans des tâches de laboratoire, en conditions de tâche simple ou double. Les résultats obtenus ont indiqué que toutes les méthodes d'évaluation de la charge sont sensibles aux conditions expérimentales. Les auteurs notent qu'avec *'Workload Profile'* l'indice global de charge (moyenne simple des huit évaluations) n'est pas le résultat le plus intéressant, mais que c'est la combinaison des différentes évaluations qui fournit le plus d'information. La *charge* est vue comme une notion intrinsèquement multidimensionnelle.

Les analyses statistiques proposées étaient basées sur l'utilisation de l'*analyse discriminante* (analyses canoniques) qui a permis de révéler des profils de charge différents dans les différentes conditions. De plus, lors de la modification des conditions d'une tâche (par exemple, accroissement d'un paramètre jouant sur la difficulté) c'est l'ensemble du profil de charge qui était modifié. La plupart des dimensions ont été impactées.

- ***Comparaison des questionnaires***

Une comparaison des propriétés psychométrique de trois questionnaires d'évaluation multidimensionnelle de la charge a été proposées par Rubio, Diaz, Martin et Puente (2004). Les questionnaires comparés étaient : *'NASA-TLX'*, *'SWAT'* et *'Workload Profile'*.

Les tâches utilisées pour la comparaison étaient une tâche de rappel (tâche de Sternberg : *« S »*) et une tâche de suivi d'une cible (tâche de *« Tracking »: « T »*), chacune étant réalisée en condition simple avec une tâche facile (S_1 et T_1) ou complexe (S_2 et T_2). Pour obtenir différents niveaux de difficulté, ces tâches ont également été combinées dans différentes conditions de double tâche (S_1T_1, S_1T_2, S_2T_1, S_2T_2).

Les résultats ont montré qu'aucun de ces questionnaires n'était intrusif (aucun effet des questionnaires sur la performance à la tâche). Ce résultat n'est pas surprenant dans la mesure où les questionnaires sont remplis après avoir terminé la tâche. En termes de sensibilité aux conditions expérimentales, les scores obtenus avec *'Workload Profile'* étaient sensibles à toutes les comparaisons, en condition simple comme en condition double. Les deux autres questionnaires étaient sensibles à des variables différentes ; et surtout, en condition de double tâche, ni *'NASA-TLX'* ni *'SWAT'* n'ont été sensibles à l'interaction entre les deux types de tâches (voir Rubio et al., 2004, pp. 71-72). Cependant, il y avait une forte validité convergente entre les trois questionnaires.

En termes de *« diagnosticité »* (*« diagnosticity »*), les auteurs souhaitaient vérifier si les évaluations obtenues permettaient de discriminer les tâches qui composaient les conditions expérimentales. Comme Tsang et Velasquez (1996), Rubio, Diaz, Martin et Puente (2004) se sont appuyés sur l'*analyse discriminante* pour répondre à cette question. Les résultats ont montré que seul 'Workload Profile' permettait de discriminer correctement les différentes conditions expérimentales. Avec ce questionnaire, deux axes étaient obtenus et tendaient à différencier (1) le type de tâche (ressources verbales ou visuelles) et (2) la complexité (condition simple ou double tâche). Les deux autres questionnaires ne permettaient de discerner qu'une variation globale de l'effort.

4.1.4 Conclusion

'Workload Profile' semble être le questionnaire le plus intéressant pour fournir des indications multidimensionnelles sur l'effort des participants à une expérience. Mais *'NASA-TLX'* est cependant le plus utilisé. Ces deux questionnaires se basent sur des approches différentes, conformément à l'opposition de Vidulich (2003) entre *'théories de la boîte noire'* et *'théories structurales'*. Avec *'NASA-TLX'* l'analyse de la tâche est prédéfinie et c'est l'opérateur lui-même qui en donne l'interprétation. Avec *'Workload Profile'*, le questionnaire est indépendant de la tâche et l'opérateur n'interprète que son effort. Dans ce cas, l'analyse de la tâche impose à l'analyste de prendre en compte d'autres indicateurs, des indicateurs objectifs, qui peuvent être confrontés aux évaluations subjectives. Ainsi, avec *'Workload Profile'* les évaluations subjectives et les analyses objectives portent sur des indications complémentaires. Avec *'NASA-TLX'*, au contraire, les évaluations subjectives (par exemple, la *'demande temporelle'*) sont redondantes avec les mesures objectives (par exemple, la *'durée de la tâche'* ou le *'nombre de mots par minute'*).

Globalement, les réserves des auteurs indiquent que l'analyse de la *charge* associée à la réalisation d'une tâche repose sur l'analyse de l'activité nécessaire pour accomplir cette tâche. La finalité de l'activité doit être prise en compte pour expliquer comment différents individus parviennent à atteindre leurs objectifs selon des profils de coût cognitif distincts.

4.2 Attribution d'un rôle causal à la charge cognitive

Comme a permis de le montrer la section précédente, la notion de *charge cognitive* est basée sur le concept opérationnel de *ressources cognitives*. Il s'agit d'estimer la part de ressources consommées au cours de l'exécution d'une tâche. Mais cette notion théorique n'est pas la finalité des études en psychologie. En psychologie ergonomique notamment, les objectifs que se fixe le psychologue consistent à étudier les déterminants de la performance atteinte par les individus (Hoc & Darses, 2007)[1]. Ainsi, les auteurs tendent à utiliser la notion de charge mentale dans cette optique d'analyse de la performance, à dominante fonctionnelle. Cette approche est abordée ici selon deux points de vue. D'abord, selon le point de vue général de la théorie de la charge cognitive, puis selon un point de vue computationnel lié à la modélisation des ressources de l'utilisateur au cours d'une tâche d'interaction téléphonique.

4.2.1 La théorie de la charge cognitive

La *théorie de la charge cognitive* a été évoquée dans le chapitre 3 au sujet des théories de l'apprentissage multimédia. Du point de vue historique, cette approche s'est développée autour des travaux de John Sweller sur les relations entre charge cognitive et apprentissage (voir Chanquoy, Tricot & Sweller, 2007; Tricot, 1998). Ces travaux ont eu une influence importante en psychologie de l'éducation et sont à l'origine d'un grand nombre d'études réalisées au cours des années 1990 et 2000 (*e.g.* Chandler & Sweller, 1991, 1996; Kalyuga, Chandler & Sweller, 1998; Paas, Renkl & Sweller, 2003), ce qui a constitué un renouveau de la notion de charge, sous l'angle de la charge cognitive. Cette théorie accorde un rôle causal à la notion de charge cognitive car elle indique que les variations des *patterns de charge cognitive* entrainent des variations de la performance d'apprentissage.

Cette théorie est classifiée ici sous le nom d'*approche fonctionnelle de la charge cognitive* car la charge cognitive y occupe un rôle causal, *i.e.* elle a une fonction dans le processus d'apprentissage.

A Présentation

L'origine de la *théorie de la charge cognitive* repose sur l'étude du processus d'apprentissage de tâches de résolution de problème (à partir de Sweller, 1976). Comme la théorie de l'apprentissage multimédia, elle répond à un objectif de conception des situations d'apprentissage. Les auteurs affirment (*e.g.* Schnotz & Kürschner, 2007, p. 471) que *« les tâches de résolution de problème sont très demandeuses en termes de capacité de mémoire de travail »*. De ce fait, les objectifs fixés ont consisté à identifier des techniques susceptibles de faciliter l'apprentissage (Sweller & Cooper, 1985; Sweller & Levine, 1982) ; et le

[1] J.M. Hoc précise à l'oral : *« La psychologie ergonomique est une psychologie de la performance »*.

développement de l'informatique a offert des perspectives à ces travaux (voir, Schnotz & Kürschner, 2007).

La théorie de la charge cognitive consiste à mettre en relation la *'structure de l'information'* avec l'*architecture cognitive'* qui permet a un apprenti de *'traiter cette information'* (Paas, Renkl & Sweller, 2003). Dans une tâche d'apprentissage, cette relation est génératrice de coût cognitif. La théorie consiste à identifier les différentes sources de coût et à les catégoriser pour en expliquer le fonctionnement dans l'architecture cognitive humaine (*Cf.* Tableau 3-2, p. 79-80). Sweller (1988 ; 1999 ; 2005) base sa théorie sur l'articulation de la *mémoire de travail* et de la *mémoire à long terme* au cours des tâches d'apprentissage, qui permet la construction de schémas en mémoire à long terme. Les articles de présentation de la théorie (*e.g.* Paas, Renkl & Sweller, 2003; Schnotz & Kürschner, 2007; Sweller, Van Merriënboer & Paas, 1998) consistent le plus souvent à décrire les différentes sources de coût cognitif, qui sont décrites succinctement dans le Tableau 4-2.

Tableau 4-2 : Les différentes sources de charge cognitive

Type de charge	Description
Charge cognitive intrinsèque *(« intrinsic »)*	Cette catégorie est nommée ainsi parce que le coût cognitif qu'elle génère est intrinsèque au matériel à apprendre. Le nombre d'éléments à apprendre et le nombre de relations entre ces éléments ont un impact sur la difficulté de l'apprentissage. Si l'interactivité entre éléments est importante, les relations entre *mémoire de travail* et *mémoire à long terme* sont plus complexes. La *charge cognitive intrinsèque* dépend du niveau d'expertise préalable de l'apprenti. Des éléments disjoints pour un novice peuvent ne constituer qu'une seule information pour un expert.
Charge cognitive inutile *(« extraneous / ineffective »)*	D'après Paas, Renkl et Sweller (2003) : *« la façon de présenter l'information aux apprentis et les activités d'apprentissage requises des apprentis peuvent aussi imposer une charge cognitive. Quand cette charge n'est pas nécessaire et qu'elle interfère avec l'acquisition et l'automatisation de schéma, elle est dite inutile ».* Ainsi, les activités qui ne sont pas directement orientées vers l'apprentissage sont vues comme *négatives* pour celui-ci et *inutiles*.
Charge cognitive utile *(« germane »)*	Comme la précédente, la charge utile dépend des activités d'apprentissage requises de l'apprenti, mais : *« Là où la charge cognitive inutile interfère avec l'apprentissage, la charge cognitive utile le facilite. »* Paas, Renkl et Sweller (2003, p. 2) La charge cognitive utile est dédiée à l'acquisition des schémas.

La Figure 4-6 propose une représentation schématique de ces trois sources de coût cognitif conforme aux descriptions proposées dans la théorie.

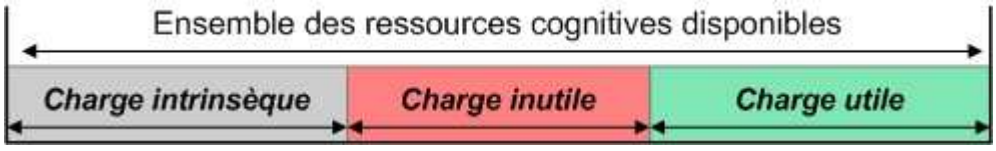


Figure 4-6 : Représentation schématique de la charge cognitive de l'individu apprenant

Cette représentation indique que les ressources cognitives existent en quantité limitée (longueur totale de la réserve de ressources). Chacune des trois sources consomme une

partie des ressources, sur un *mode additif* (Schnotz & Kürschner, 2007; Sweller, 2005; Tindall-Ford, Chandler & Sweller, 1997). Ainsi, les ressources consommées du fait de l'une de ces sources de coût cognitif ne peuvent être allouées à un autre type de traitement.

Deux phénomènes principaux sont générateurs de charge cognitive inutile : (1) les phénomènes liés à la redondance des informations et (2) les phénomènes liés au partage de l'attention. Chacun de ces deux types a été commenté dans les pages qui précèdent.

B Justification du rôle de la charge cognitive pour l'apprentissage

Dans l'article qui fonde la théorie de la charge cognitive (Sweller, 1988) l'auteur distingue deux types d'activité au cours de l'apprentissage :

(1) Une activité d'*atteinte du but* et ;

(2) Une activité d'*acquisition de schémas.*

Il indique que :

> *« L'individu dont la capacité de traitement cognitif est entièrement dévouée à l'atteinte du but porte son attention sur cet aspect du problème à l'exclusion des caractéristiques du problème nécessaires à l'acquisition de schéma. » (p. 262, traduction libre)*

Ainsi, des phénomènes d'interférence existent entre les deux types de traitements. Pour en étudier les déterminants, l'auteur propose de quantifier la charge cognitive associée à une tâche de résolution de problème. Pour faire cette quantification, il utilise une méthode de décomposition des stratégies des participants dans une tâche de résolution de problème, selon que la stratégie employée est une *stratégie 'moyen-fin'* ou une *'stratégie sans but spécifique'*. Cette décomposition consiste en une formalisation des états successifs rencontrés par l'individu qui réalise la tâche. Elle représente les opérations que doit faire l'individu, par exemple, ajouter une information ou un sous-but en mémoire de travail, résoudre une addition, vérifier si l'état final est atteint, etc. Il s'agit donc d'une analyse fonctionnelle du processus étudié, formalisée sous la forme d'un modèle computationnel. L'auteur utilise ces analyses pour proposer une méthode de quantification de la charge cognitive. Elle est évaluée par : (1) le nombre d'*éléments à maintenir en mémoire*, (2) le nombre de *règles de production*, (3) le nombre de *cycles nécessaires à l'atteinte de la solution* et (4) le nombre total de *conditions qui doivent être associées*.

A partir de ces éléments, l'auteur présente une expérience basée sur la résolution de problèmes trigonométriques. Les participants utilisant une *stratégie moyen-fin* ont obtenu de moins bonnes performances de compréhension que ceux qui n'avaient *pas de but spécifique*. D'après ces résultats, l'auteur propose les conclusions suivantes (Sweller, 1988, p. 284) :

(1) Les mécanismes liés aux deux types d'activité imposent une charge cognitive lourde ;

(2) Ces mécanismes sont *substantiellement distincts* ;

(3) L'effort requis pour résoudre le problème peut ne pas assister l'acquisition de schéma ;

(4) L'acquisition de schéma étant possiblement le composant le plus important du développement de l'expertise, celui-ci peut être retardé par la focalisation sur la résolution de problème et ;

(5) Les théories indiquant que la résolution de problème permet l'apprentissage doivent être modifiées en conséquence.

Cette étude indique que la stratégie de recherche mise en place par un individu peut produire un effet sur la compréhension qu'il a du problème. L'auteur en déduit qu'il existe une opposition entre les deux mécanismes identifiés. Ils entreraient en concurrence pour les ressources. Pour la *théorie de la charge cognitive*, cette distinction a permis de valider l'opposition entre *charge cognitive intrinsèque* et *charge cognitive inutile*. La prise en compte d'une *charge cognitive utile* est apparue plus tard (Sweller, 1999).

C Conclusion

Dans l'analyse fonctionnelle proposée par Sweller (1988), tous les éléments décrits sont relatifs à l'activité cognitive de l'individu. Les indications fournies sur la charge cognitive sont toutes issues du comptage des éléments de cette description formelle. La notion de *charge cognitive* n'apparaît pas dans cette formalisation. Elle est induite par l'auteur après l'analyse. Strictement, aucune *« charge cognitive »* n'est décrite. Cette notion n'est d'ailleurs pas définie dans l'article. L'auteur l'utilise pour apporter une indication générale sur les stratégies décrites. Il serait tout aussi exact de parler de *« complexité de la stratégie »*. De ce fait, la description quantitative qui consiste à évaluer la charge cognitive est une réduction de la richesse de l'analyse proposée. Mais l'utilisation de cette notion générale a cependant permis à cette théorie de connaître un retentissement important.

Par ailleurs, une remarque peut être faite sur la notion de *charge cognitive inutile* (*« extraneous/ineffective cognitive load »*). Le choix de ce terme provient de la finalité de la théorie : l'apprentissage. Il est important de remarquer que ce positionnement théorique induit un préjugé négatif envers ce type de traitement ; préjugé impropre à l'analyse.

4.2.2 Le problème de la charge cognitive en DHM

Le type de modélisation utilisé par Sweller (1988) repose sur le même type de descriptions que celles utilisées en Intelligence Artificielle. Les techniques de formalisation des tâches sont communes avec celles utilisées en DHM (*Cf.* Chapitre 2) et en *« psychologie computationnelle »* (*Cf.* Ce chapitre, partie précédente). Ces techniques permettent de décomposer les procédures mises en place par les participants au cours de l'exécution des tâches étudiées. Elles semblent plus propices à l'analyse que la notion générale de charge cognitive. Des techniques du même type ont été utilisées dans des travaux en DHM, pour évaluer l'impact des surcharges sur la production langagière et pour évaluer le niveau de charge de la mémoire au cours d'une interaction avec un service téléphonique.

Certains travaux ont eu pour objectif d'identifier les surcharges grâce à la détection des modifications comportementales qu'elles induisent. Certains cherchent à en évaluer les effets dans le dialogue, dans une optique de conception (*e.g.* Le Bigot, Rouet & Jamet, 2007; Oviatt, Coulston & Lunsford, 2004). D'autres cherchent à concevoir des modèles computationnels utilisant un ensemble d'indicateurs (comportementaux ou vocaux) pour détecter les surcharges en temps réel (*e.g.* Baber, Mellor, Graham, Noyes, & Tunley, 1996; Berthold & Jameson, 1999; Jameson et al., 2006; Müller, Großmann-Hutter, Jameson, Rummer, & Wittig, 2001; Wilamowitz-Moellendorff, Müller, Jameson, Brandherm, & Schwartz, 2005; Yin & Fang, 2007). Ce type de modèle peut permettre d'automatiser certaines procédures adaptatives.

A Indicateurs de surcharge dans le dialogue

L'article de Baber, Mellor, Graham, Noyes et Tunley (1996) permet par exemple de montrer que des conditions imposant une *charge de travail* importante à l'utilisateur peuvent avoir un effet sur la parole qu'il produit en retour. Dans une première expérience, ils ont demandé aux participants de lire des plaques minéralogiques à l'aide de l'alphabet ICAO (alphabet aéronautique : *« Alpha, Bravo, Charlie, Delta, etc. »*) à la fréquence, soit de 20 plaques à la minutes, soit de 33. Les résultats ont montré que dans le second cas la charge de travail (évaluée avec le questionnaire NASA-TLX) était plus importante et que, par ailleurs, la performance verbale se dégradait. En effet, l'utilisation des enregistrements pour une reconnaissance avec un logiciel de reconnaissance vocale a donné lieu à une performance moins bonne pour la fréquence de 33 plaques par minute. Dans une seconde expérience, les participants étaient mis dans des conditions de double et triple tâche : avec une *'tâche d'identification de cible'* et une *'tâche de calcul arithmétique'*. Là aussi les évaluations subjectives de la charge de travail ont été sensibles, mais la performance de reconnaissance vocale n'était que faiblement dégradée. Les auteurs indiquent que les modifications de la parole portaient surtout sur les espacements entre les mots et sur le découpage du rythme d'émission des syllabes. Cette fois encore, les participants ont adapté leur comportement aux conditions de la tâche. Les auteurs en ont conclu que la charge de travail est un sujet important dans la recherche sur la reconnaissance vocale car elle impacte la parole.

Cet article est représentatif. Les mises en situation présentées sont très peu écologiques. Elles sont destinées à prouver l'impact de la charge de travail. En termes de *définition*, les auteurs se réfèrent d'abord à la notion de ressources vue comme un *'pool de ressources'* unique (Norman et Bobrow, 1975), puis au *modèle de ressources multiples* (Wickens, 1984), puis ils font référence aux ressources nécessaires à la production de la parole sur la base des théories linguistiques classiques (niveaux sémantique, syntaxique, lexical et articulatoire). Toutes ces définitions sont cognitives, même si elles diffèrent les unes des autres. Ces différents points ne sont pas réellement articulés entre eux. Ils sont présentés pour montrer l'importance de la notion. Du *point de vue méthodologique*, les auteurs s'en remettent au questionnaire NASA-TLX qui relève d'un modèle théorique différent (voir plus haut). Ce choix n'est pas justifié dans l'article et on est en droit de se demander comment les auteurs lient

tous ces éléments entre eux. Pris globalement, le raisonnement est le suivant : *la charge de travail est importante, donc elle est importante*. En effet, partant de préconceptions théoriques qui leur font préjuger de l'importance de la notion, les auteurs se donnent les moyens méthodologiques qui leur permettent de renforcer cette idée. Ce raisonnement est cyclique et il passe, de plus, par des mises en situation peu écologiques. Son intérêt consiste à identifier les modifications qui apparaissent quand la tâche est plus complexe.

Des raisonnements et des méthodologies équivalents peuvent être identifiés dans la plupart des articles sur ce thème. Les auteurs eux-mêmes formulent parfois des réserves du même type. Par exemple, Jameson et al. (2006) indiquent dans leur conclusion :

> *« Il peut être supposé que les probabilités spécifiques d'une reconnaissance correcte (des surcharges) dépendent des caractéristiques particulières à la situation (…). Pour nos analyses, il était certainement utile que la situation expérimentale soit hautement contrainte » (Traduction libre).*

- ***Indices comportementaux issus de la parole de l'utilisateur***

Au-delà du rôle attribué à la notion de charge de travail, l'objectif est de mettre en place des modèles computationnels basés sur l'analyse des caractéristiques de la parole pour identifier les situations dans lesquelles l'utilisateur est perturbé. Ces caractéristiques sont diverses et ne peuvent être interprétées isolément, ce qui impose d'utiliser des techniques permettant d'assumer cette complexité, telles que la modélisation par *réseaux bayésiens*. Par exemple, Jameson et al. (2006) se basent sur :

(1) Le *nombre de syllabes* dans l'énoncé ;

(2) Le *taux d'articulation*, en nombre de syllabes par seconde ;

(3) Les *temps de pauses silencieuses* ;

(4) Les *temps de pauses continues*, par exemple, suite à un « euh… » ;

(5) La *fréquence des hésitations courtes* (inférieures à 200 ms) ;

(6) Le *temps entre stimulus et prise de parole*, et ;

(7) Les *disfluences* (disjonctions dans les énoncés (reprises syntaxiques, etc.)

Quelles que soient les explications fournies[1], cette démarche est utile pour la détection de perturbations dans la parole des utilisateurs. Cependant, Jameson et al. (2006) remarquent que pour atteindre cet objectif, il est nécessaire de caractériser les situations liées au déroulement de la tâche pour interpréter les symptômes de perturbation de la parole. En elle même, l'utilisation d'un modèle computationnel n'est pas suffisante.

[1] Par exemple, la première phrase de Berthold et Jameson (1999), utilisée pour justifier leur approche, est la suivante : *« Quand des cosmonautes dans la station spatiale Mir communiquent avec le centre de contrôle, leur parole est surveillée par des psychologues pour en analyser les symptômes de stress. »* (Traduction libre.) Bien que ce fait ne soit pas en doute, il ne prouve aucunement le rôle causal, ni du stress, ni de la charge cognitive, dans l'utilisation d'un système de DHM…

B Dépassement de la mémoire de travail en interaction téléphonique ?

Une étude a été consacrée plus spécifiquement au rôle de la *mémoire de travail* dans le DHM (Huguenard, Lerch, Junker, Patz, & Kass, 1997). L'objectif de ces auteurs était de proposer une modélisation computationnelle de la mémoire de travail qui distingue les processus de *'traitement'* (*« processing »*) et de *'stockage'* (*« storing »*) des informations. Ils notent que l'interaction téléphonique (PBI pour *« phone based interaction »*) *« est un domaine de tâche qui a des avantages considérables pour la conception de modèles cognitifs des défaillances de la mémoire de travail et pour l'examen empirique des erreurs comportementales dues aux limites de la mémoire de travail. »* (p. 71, traduction libre).

La modélisation utilisée était basée sur le modèles 3CAPS (Just & Carpenter, 1992). Trois types de facteurs pouvant impacter la mémoire de travail étaient pris en compte :

(1) La *structure des menus* (en *largeur* et en *profondeur*). Deux structures comparées :

- ***'PBI-deep'*** : Profonde et peu large (Quatre niveaux de trois choix : ***3*3*3*3***) ;
- ***'PBI-broad'*** : Large et peu profonde (Deux niveaux de neuf choix : ***9*9***).

(2) La *capacité de mémoire de travail*, qui est spécifique à chaque individu (évaluée par le questionnaire de Daneman et Carpenter, 1980) et ;

(3) Les *caractéristiques de la tâche* (complexité de la consigne). L'application choisie proposait des menus en DTMF (sélection par les touches du téléphone).

Les hypothèses portaient sur les taux d'erreur au cours des dialogues, soit des *'pertes d'information'* qui imposent de revenir sur un menu (réécoute immédiate), soit des *'erreurs de navigation'* qui imposent des retours en arrière (suite à des choix erronés), soit l'*échec de la tâche'* par application d'une mauvaise consigne ou par échec dans la réalisation.

Les résultats sur les taux d'échecs ont montré une différence globale entre les deux structures de menus qui était due aux *erreurs de navigation* (17,73% pour *'PBI-deep'*, contre 7,74% pour *'PBI-broad'*). Les autres taux d'erreur étaient similaires entre les deux versions. Par ailleurs, les *capacités de mémoire de travail* des participants (classés en forts, moyens et faibles) n'ont pas eu d'impact sur les *pertes d'information* ni sur les *erreurs de navigation*, mais elles ont eu un effet sur les *échecs de la tâche* (de forts à faibles : 5,73%, 11,19% et 12,59%). Par ailleurs, la complexité de la tâche a également eu un impact, passant d'un taux d'échec de 3,83% pour la consigne la plus simple à 13,28% pour la plus complexe.

Les auteurs ont réalisé diverses analyses statistiques afin d'estimer l'impact relatif de chacune des variables. Ils en concluent que la structure de dialogue profonde (*'PBI-deep'*) ne génère pas une charge de traitement supplémentaire. Ils indiquent que le faible impact des capacités de mémoire des participants est lié au protocole utilisé et qu'il est possible de concevoir une tâche qui surchargerait plus et donnerait lieu à des *pertes d'information* plus fréquentes. Pour appuyer leurs explications, ils proposent un graphique (Figure 4-7), qui met en relation : capacité de mémoire, charge en mémoire et performance à la tâche.

D'après ce graphique, la performance optimum à la tâche est obtenue pour une charge en mémoire raisonnable, qui se trouve en deçà du seuil des capacités des participants, importantes ou non. Les auteurs notent que leurs conditions expérimentales restaient en majorité en deçà du seuil. Les structures de menu proposées avaient pourtant une complexité importante (l'arbre conduisait à 81 feuilles). Mais bien que celle-ci ne dépasse pas les capacités des utilisateurs, de nombreux problèmes d'ergonomie se posent dans cette structure. Pourquoi alors en appeler nécessairement à la surcharge de la mémoire ? La cognition humaine n'a-t-elle pas d'autres propriétés tout aussi importantes ? Les auteurs indiquent notamment que les différences entre stratégies proviennent des conséquences spécifiques à certains choix, surtout en cas d'erreur. Ces conséquences dépendent de la structure de menu et sont relativement indépendantes des capacités mémorielles. Ils renvoient ainsi au processus en jeu plutôt qu'aux capacités sur lesquelles il s'appuie. Dans ce sens, la question qui se pose pour le psychologue consiste à savoir comment l'utilisateur interprète cette structure dans la limite de ses capacités, quelle que soit cette limite.

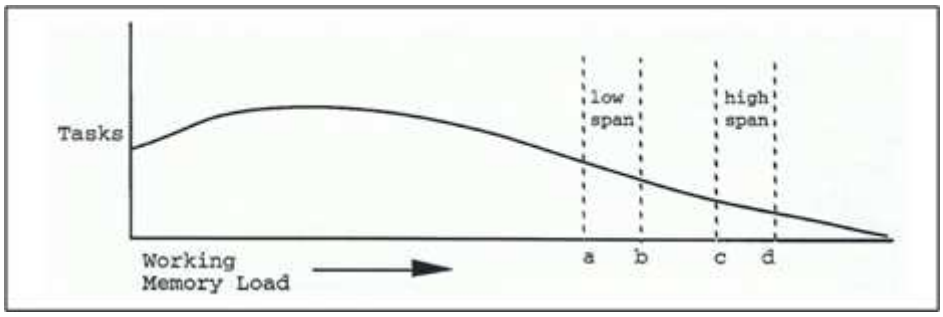


Figure 4-7 : Relation entre charge en mémoire de travail et performance à la tâche

Finalement, les auteurs notent :

> *« Par conséquent, la structure profonde ('PBI-deep') requiert théoriquement un traitement supérieur par rapport à la structure large ('PBI-broad'), mais en fait l'effort de traitement effectivement accompli est inférieur à l'effort attendu. » (p. 97, trad. libre)*

Enfin, ils concluent par deux implications pour le champ des interactions homme machine :

- *« Premièrement, la phrase ''surcharge informationnelle'' (« information overload »)* est souvent discutée comme un problème majeur pour l'interaction homme machine, *mais ce qui est surchargé est rarement clair. »* (p. 97, traduction libre.)
- *« Deuxièmement, nous avons montré que l'utilisation d'une modélisation cognitive computationnelle fournit une approche systématique pour conduire des recherches en interaction homme machine »* (p. 98, traduction libre.)

Ainsi, ces auteurs semblent douter de la pertinence théorique de la notion de *surcharge informationnelle* et ils indiquent que la modélisation cognitive offre une méthodologie utile pour aller plus loin dans l'analyse.

4.3 Conclusion

Les travaux qui accordent à la charge cognitive un statut explicatif intermédiaire sont basés sur l'idée d'une *'circulation des informations*' héritée des prémisses présentées en début de chapitre et relevant de la métaphore informatique pour expliquer le fonctionnement cognitif humain. Or, ces approches fonctionnelles de la charge cognitive renvoient à la nécessité de formaliser l'activité de l'individu pour analyser ses performances.

Pourtant, comme cela a été indiqué à plusieurs reprises dans les premiers chapitres, le processus interactif (ou '*processus de grounding*' dans les termes de Clark) qui permet à des partenaires d'entrer en interaction est premier et les processus de « haut niveau » en dépendent (*principe de subsomption*, voir Brooks, 1991). Ils sont inscrits en lui et non l'inverse. Ainsi, le statut privilégié accordé aux notions de *'traitement de l'information'* et de *'charge cognitive'* est le fruit d'une exagération du rôle de ces processus. Les études qui se focalisent sur ces traitements et les théories qui leur attribuent un rôle explicatif central pourraient bien passer, en réalité, à coté du problème majeur qui se pose dans l'interaction : l'*organisation des comportements qui la constituent* (l'interaction). Si c'est le cas, il serait alors raisonnable de penser que les concepteurs d'un système interactif doivent, en premier lieu, chercher à faciliter le processus interactif pour offrir aux utilisateurs un outil acceptable ; et dans un second temps seulement – et sur cette base – se préoccuper des traitements cognitifs et des activités d'apprentissage des contenus. Cette remarque est importante car elle incite à emboîter correctement les objectifs de conception.

Les bases théoriques qui fondent la notion de *charge cognitive* sont assez largement remises en cause aujourd'hui. Par exemple, les travaux sur la *mémoire de travail à long terme* (MDT-LT, Ericsson & Kintsch, 1995) ont relativisé l'isolement d'un processus central de traitement des informations. De même, les modèles computationnels (Anderson et al., 2004 ; Kieras et Meyer, 1997 ; 2000) tendent à élargir la problématique aux processus de contrôle de l'activité. Dans ce sens, Hoc et Amalberti (2007) expliquent que les études portent un intérêt croissant au rôle des automatismes pour contrôler l'activité. Par ailleurs, on trouve également des critiques plus radicales et plus générales (Glenberg, 1997; Varela, Thompson & Rosch, 1993) qui soulignent l'importance du corps et de l'orientation dans l'action.

La partie expérimentale de la thèse devrait permettre de fournir des résultats allant dans le sens de ces critiques et orientés vers l'analyse des actions dans des dialogues finalisés avec un système de DHM simulé (*Magicien d'Oz*).

PARTIE EXPERIMENTALE

Chapitre 5 Analyse et Problématique

L'objectif de la thèse est de fournir une *évaluation expérimentale* des effets produits par des modifications des *énoncés du système* dans un contexte de DHM. Ce chapitre va d'abord permettre de proposer une analyse des services dialogiques utilisés dans les expériences et d'en déduire une problématique plus précise.

5.1 Analyse des services dialogiques

L'analyse des services est basée sur une synthèse des éléments théoriques présentés dans les chapitres précédents. Cette synthèse est ensuite mise en relation avec les descriptions individuelles des services dialogiques utilisés dans les expériences de la thèse.

5.1.1 Schéma de synthèse de la situation de communication

La Figure 5-1 a été conçue pour donner une vision générale de la situation de communication entre un système de dialogue et un utilisateur. Sa description préalable va permettre de dégager le jeu de relations interne à une communication.

Dans ce schéma, la structure hiérarchique et temporelle de l'interaction (*e.g.* Roulet, 1981; Roulet et al., 1985, chapitre 1) est en arrière-plan. Cette précision indique que tout énoncé produit par l'un des partenaires en communication joue un rôle (ou un ensemble de rôles) dans cette structure hiérarchique. L'énoncé débute par une prise de parole et se termine par une cession de parole (Sacks, Schegloff & Jefferson, 1974). Le locuteur prend part à l'échange, qu'il contribue à structurer. L'énoncé est ainsi situé au cœur de l'interaction (ici, *'interaction'* est synonyme de *'conversation'*). Les notions de *contact*, de *perception*, de *compréhension*, d'*attitude* et d'*émotion* (Allwood, Traum & Jokinen, 2000, *Cf.* Chapitre 1 - Le point de vue d'Allwood) ont été placées relativement à cette structure. Cette structure hiérarchique d'arrière plan indique comment la conversation est organisée. – La structure des dialogues expérimentaux est présentée plus loin. –

Le premier plan représente la situation de communication *dans laquelle* les interlocuteurs sont liés. Bien qu'elle puisse être envisagée à chacun des niveaux (ou rangs) hiérarchiques, cette situation est rapportée ici en priorité au niveau de l'*intervention* (ou *« énoncé »*). C'est pourquoi l'effet de perspective est dessiné à partir de ce niveau. C'est également sur ce niveau que s'est appuyé Austin (1962) pour avancer son analyse et il est naturel de retrouver les trois dimensions de l'acte mises en évidence par cet auteur (*'locutoire'*, *'illocutoire'* et *'perlocutoire'*). De ce fait, le schéma proposé ici est construit comme le schéma qui a été

proposé pour synthétiser l'analyse d'Austin, présentée au début du chapitre 1 (*Cf.* Figure 1-1). Les expressions utilisées dans la Figure 5-1 (*« marqueurs locutoires »*, *« conventions illocutoires »* et *« effets perlocutoires »*) ont été mises au pluriel pour signifier la pluralité des fonctions en jeu dans la situation.

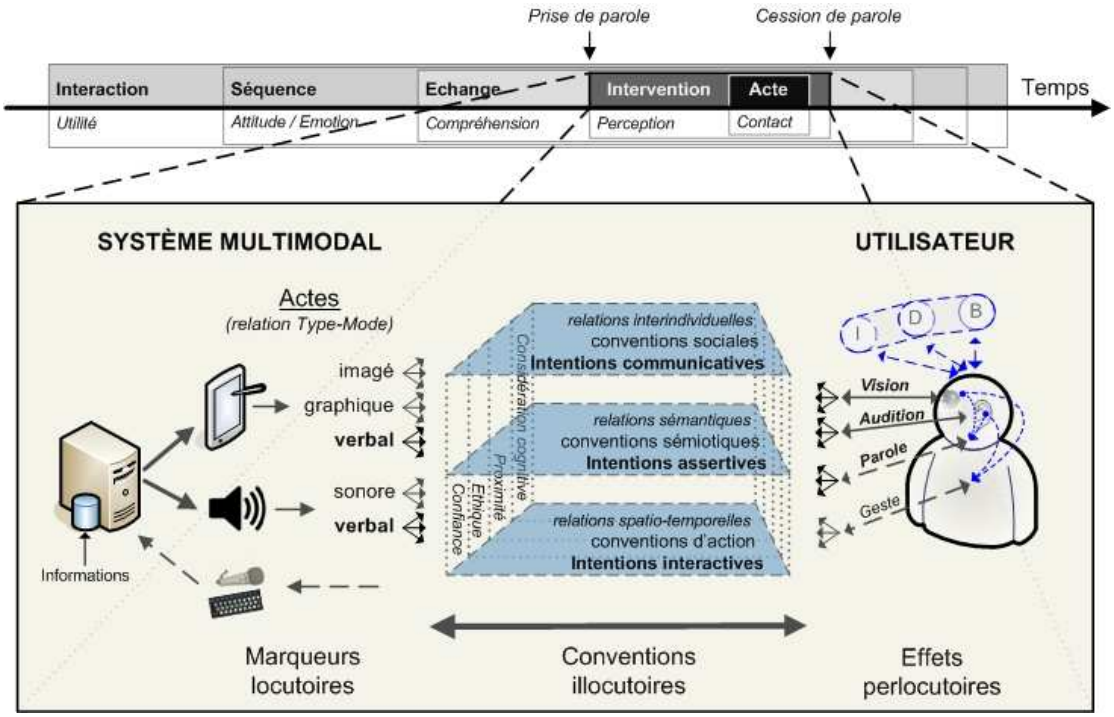


Figure 5-1 : Schématisation de la situation lors de la production d'un énoncé système

A Les marqueurs locutoires

Dans ce schéma, le système informatique est le producteur des *'marqueurs locutoires'*. La notion de *'marqueur locutoire'* est substituée à celle de *'signification locutoire'* utilisée dans le chapitre 1, dans la présentation des travaux d'Austin (1962). Le terme « marqueur » est emprunté à Wilson (1997). Il indique que la signification n'est pas le fait du locuteur, mais qu'elle est liée à l'interprétation par un individu en position d'observateur, qui perçoit l'énoncé.

On peut remarquer que l'information est puisée dans le disque dur du système (ou issue des processus inférentiel du système). Cette remarque indique que la notion d'*information* relève d'une réalité technique liée au stockage et au transport des « *informations* » dans les « systèmes *informatiques* ». Ici, l'énoncé n'est pas vu comme un ensemble d'*informations*, mais comme un ensemble d'*actes*. A ce niveau, les entrées du système (*clavier, micro, souris*, en bas à gauche) sont représentées pour indiquer l'aspect cyclique du phénomène de communication.

Les sorties sont scindées entre périphériques visuel (*écran*) et auditif (*haut-parleur*). Le *code* utilisé pour présenter l'*information* met en relation un certain *type de contenu* et un certain *mode de présentation* (la *modalité*), pour produire les effets voulus dans l'interaction :

- Le *type de contenu* renvoie à une catégorisation des *actes,* ou des *informations*, selon le point de vue ;
- La *modalité de présentation* renvoie quant à elle au choix combiné d'un dispositif de sortie (écran, haut-parleur, etc.) et d'un langage d'interaction (verbal, graphique ou pictural) comme indiqué par Nigay et Coutaz (*e.g.* 1994).

Dans la suite du document, l'expression *« relation type-mode »* sera utilisée. Cette expression sera mieux expliquée après la présentation des résultats de l'expérience 3 et dans la problématique de l'expérience 4. Cette relation implique que, pour chaque '*unité de contenu'* (*information* ou *acte*), le choix d'un certain *mode* de présentation produira des effets spécifiques sur l'interlocuteur et, globalement, sur la performance à la tâche. Les modalités indiquées (*'pictural'*, *'graphique'*, *'sonore'* et *'verbal'*) sont sensibles à cette relation. Elles produisent des effets (perlocutoires) sur l'interlocuteur (l'*utilisateur*) par l'intermédiaire des conventions illocutoires connues et utilisées par cet interlocuteur. Les triples flèches issues des différentes *modalités* indiquent que tout acte peut être interprété en parallèle selon, au moins, ces trois différents plans (présentés ci-dessous).

B Les conventions illocutoires

Comme indiqué par Austin (1962), tout contenu adressé à un *destinataire* est interprété par lui au travers de *'conventions illocutoires'*. A ce niveau, on peut distinguer plusieurs plans d'interprétation. Tout acte peut faire l'objet d'interprétations dans chacun des trois plans proposés. Ces interprétations sont des *intentions attribuées au locuteur par le destinataire*. Elles peuvent correspondre aux intentions d'origine du locuteur. Les trois plans proposés sont issus de la littérature.

Tout d'abord, l'article de Clark et Carlson (1982) indique qu'il est nécessaire de distinguer l'intention *« communicative »* de l'intention *« assertive »,* qui forment les deux plans représentés en haut et au milieu. Plus bas dans le schéma, la distinction entre *intention assertive*, au milieu, et *intention interactive*, en bas, est moins consensuelle. Par exemple, Roulet (1981) distingue la *« fonction interactive »* de la *« fonction illocutoire »* pour analyser les relations entre les actes au sein de l'énoncé (*Cf.* Tableau 1-6, p. 18). La proposition faite ici est un peu différente[1]. Elle est inspirée des remarques de Goffman (1974; 1987) qui insiste sur la composition de la *parole* en un ensemble *« mot-geste »*. Ces remarques sont également présentes chez Allwood, Traum et Jokinen (2000). La parole a une dimension gestuelle qui est, elle aussi, intentionnelle. Lors de l'interprétation, chaque plan apporte des informations complémentaires qui enrichissent le sens dégagé par l'interprète. Cette distinction de trois plans correspond à la triple coprésence de Clark et Marshall (1981).

Dans le schéma, les trois plans sont superposés verticalement. Ils ne doivent pas pour autant être interprétés comme des *niveaux d'abstraction* (*e.g.* Rasmussen, Pejtersen & Goodstein,

[1] Au sujet des difficultés terminologiques liées aux préférences des auteurs, voir Kerbrat-Orecchioni (1995).

1994). En fait, dans chacun des trois plans, différents niveaux d'abstraction peuvent être envisagés. Ces plans distinguent des aspects différents de la réalité, illustrés par des types différents de relations (*interindividuelles*, *sémiotiques* et/ou *spatio-temporelles*). Et bien sur, ces différents aspects entretiennent entre eux des relations. La formalisation parallèle de ces trois plans dans un système artificiel permettrait d'étudier ces relations plus en détail.

Un autre aspect vient encore compléter cette description. Les caractéristiques coopératives proposées par Allwood et ses collaborateurs (*i.e.* considération cognitive, proximité, éthique et confiance, *Cf.* Allwood, 1995; Allwood, Traum & Jokinen, 2000) viennent entrecouper ces plans. Elles permettent de qualifier les relations entre les interlocuteurs. C'est pourquoi elles sont représentées, dans ce schéma, au niveau des conventions illocutoires. Il doit être entendu qu'au-delà de cette représentation schématique, ces qualificatifs ont une portée large et qu'ils peuvent permettre de *qualifier l'équipe* en communication dans son ensemble tout autant que l'attitude de chacun des interlocuteurs.

C *Les effets perlocutoires*

Enfin, l'énoncé produit des *'effets perlocutoires'* chez le destinataire. Des effets divers sont produits en parallèle et peuvent être identifiés relativement à chaque plan d'intention :

- Sur le *plan interactif*, les effets se déploieront dans l'espace et dans le temps ;
- Sur le *plan assertif*, le destinataire acceptera plus ou moins le contenu asserté ;
- Sur le *plan communicatif*, l'énoncé positionnera socialement les interlocuteurs et éventuellement d'autres individus.

Chez le destinataire, l'énoncé est perçu par les organes sensoriels dont la vision et l'audition sont les principaux représentants (*Cf.* Wickens, 1984). A partir de ces perceptions, des *couples fonctionnels* (Rasmussen, Pejtersen & Goodstein, 1994) sont mis en jeu pour générer des réponses. On peut également parler de *boucles perception-action* (Garrod & Pickering, 2004) qui mettent en jeu les *affordances* (Gibson, 1979; Norman, 1999) dans le contexte. D'une façon générale, ces termes désignent la capacité des êtres biologiques à produire des séquences d'action organisées face à des stimuli variés.

La production de ces séquences d'action se base, d'une part, sur l'organisation physique du corps qui les produit (système nerveux, arcs réflexes, etc.) et, d'autre part, sur l'expérience de ce corps acquise par répétition tout au long de son existence (automatisation et apprentissage). Wickens (2002, *Cf.* chapitre 4) identifie quatre boucles perception-action principales : (1) *vision-geste*, (2) *vision-parole*, (3) *audition-geste*, et (4) *audition-parole*. Il est possible de prédire des *'temps de réaction'* et des *'niveaux de précisions'* différents sur ces différentes boucles. Par exemple, la loi de Fitts (1954) porte sur la boucle (1) *vision-geste*[1].

[1] La loi de Fitts établit le temps nécessaire au pointage d'une cible avec une souris en fonction de la taille de la cible et de la distance à parcourir (formule : Temps = a + b* $\log_2$ ((Distance à la cible / Largeur de la cible) + 1) où a et b sont des constantes).

Chacune de ces boucles d'action peut être plus ou moins adaptée au traitement des divers stimuli de l'environnement et au type de réponse attendue.

Par ailleurs, ces réponses sont produites par un sujet doué de connaissances et qui se positionne (dans chacun des plans) par rapport à l'information reçue. Dans le schéma de la Figure 5-1, ces connaissances sont évoquées par les lettres BDI, en référence aux systèmes de dialogue et à l'approche agent (Bratman, Israel & Pollack, 1988; Sadek, Bretier & Panaget, 1997). D'autres approches sont évidemment possibles, la plus fréquente en psychologie cognitive étant issue de la *théorie des schémas* (Johnson-Laird, 1980). Pour acquérir (par construction) des informations sur l'état des connaissances du destinataire à partir de ses énoncés, le système doit se baser sur les indices observables à sa disposition (Qu'écoute-t-il ? Que regarde-t-il ? Quelles sont ses réponses ? Comment les organise-t-il ?) et sur ses propres connaissances à ce sujet, acquises préalablement.

D Conclusion

On voit à travers ce schéma que la situation de communication n'est pas une chose simple. Il en va de même de l'interprétation d'un énoncé qui, pour être complète, doit prendre en compte tous ces aspects. C'est loin d'être le cas aujourd'hui dans la plupart des travaux expérimentaux qui (1) ne prennent bien souvent en compte que l'intention assertive et qui ignorent ou rejettent la dimension comportementale du sujet apprenant (comme c'est le cas dans les travaux de Mayer), qui (2) ne considèrent pas la *'relation type-mode'* pour présenter les informations (*e.g.* travaux sur la *redondance* des informations) et qui (3) ne s'intéressent souvent qu'à une seule boucle perception-action sans mentionner son caractère spécifique (ce dont l'isolement, au moins technique, des communautés *'dialogue'* et *'IHM'* est à la fois le symptôme et la cause).

Enfin, une remarque peut être ajoutée au sujet de la notion d'*interaction*. On peut remarquer que cette notion apparaît à deux positions dans le schéma. Il s'agit, d'une part, de l'unité la plus grande du modèle hiérarchique qui se situe en arrière plan, et d'autre part, de l'un des plans intentionnels qui constituent les conventions illocutoires. Cela montre l'importance de la notion. Il s'agit à la fois du bain dans lequel les interlocuteurs communiquent, qui peut être plus ou moins structuré et faire l'objet de connaissances et de croyances (historique du dialogue), et d'un plan d'intention lié aux actes produits en temps réel, qui peut faire l'objet d'interprétations spécifiques de la part du destinataire.

5.1.2 Structure hiérarchique des services dialogiques

La synthèse qui vient d'être proposée décrit le système relationnel en jeu à chaque instant dans le DHM. Le premier plan de la Figure 5-1 a été détaillé dans le commentaire. Cette partie permet maintenant de revenir sur la *structure hiérarchique* d'arrière-plan telle qu'elle peut être décrite dans le cadre de l'utilisation d'un système de DHM.

Tous les services dialogiques ont une organisation comparable. Les méthodes de conception sont relativement similaires d'un service à l'autre et doivent, de plus, être adaptées à l'organisation naturelle des conversations.

Un service peut être décomposé sous la forme d'une succession d'états (*Cf.* chapitre 2). Chacun de ces états peut également être décomposé en actes élémentaires qui peuvent être observés à différents niveaux (*Cf.* chapitre 1). Pour décrire cette décomposition, la paire adjacente question-réponse est le niveau d'observation le plus saillant (*Cf.* notamment, Sacks, Schegloff et Jefferson, 1974, voir chapitre 1). Les services dialogiques existants sont généralement composés d'un faible nombre de paires adjacentes différentes. Une structure à deux niveaux permet de les décrire.

A Structure locale : la « Phase » de dialogue

Comme le notent Bangerter, Clark et Katz (2004) dans la conversation téléphonique entre humains, tout dialogue est organisé en trois phases principales : (1) une *ouverture*, (2) un *corps* et (3) une *clôture*. De même, un service de DHM est structuré en plusieurs phases. Ces phases correspondent à des états du système au cours desquels la tâche à réaliser connaît un certain niveau d'avancement. Au cours d'une *phase*, le système délivre un contenu à l'utilisateur et attend une *commande* (ou une *demande*) qui lui permettra d'évoluer vers un autre état.

- ***Organisation d'une « phase de dialogue »***

Chaque *phase* peut être décomposée à un niveau plus fin. Elle correspond, au minimum, à une *paire adjacente question-réponse*, c'est-à-dire à un *tour de parole* de la part du système auquel s'ajoute un *tour de parole* de la part de l'utilisateur. C'est à ce niveau que se pose le problème des *prises* et des *cessions* de parole (Sacks, Schegloff et Jefferson, 1974).

Dans le dialogue avec un service automatique, le caractère artificiel de la communication impose de mettre en place des comportements plus explicites que dans le dialogue entre humains. Le système propose un énoncé constitué des trois types de contenus nécessaires au guidage de l'utilisateur :

- L'*écho'* permet de répéter la demande formulée par l'utilisateur ;
- La *'réponse'* présente les contenus qui correspondent à cette demande ;
- La *'relance'* indique à l'utilisateur qu'une nouvelle demande est attendue.

Cette structure peut être utilisée pour qualifier la performance. Le comportement de l'utilisateur peut être décrit relativement à celui du système, utilisé comme repère.

B Structure globale : le « Dialogue »

L'utilisateur appelle le service avec un certain objectif. Pour atteindre cet objectif, il réalisera un parcours parmi les phases qui composent le service. Ce parcours correspond au

« dialogue » de l'utilisateur avec le système. L'ensemble des transitions d'états accomplies au cours d'un dialogue permettent de décrire la structure globale de la conversation. Cette structure peut également être utilisée pour qualifier l'efficacité du dialogue (franchissement des phases, retours en arrière, etc.)

- ***Organisation d'un « dialogue »***

Dans un dialogue, différents allers-retours peuvent être nécessaires entre les phases ; et à différents passages dans la même phase, les buts peuvent être différents. Par exemple, la question : « Que désirez-vous ? » n'a pas le même sens s'il s'agit d'une requête initiale ou de la correction d'une incompréhension. Pour différencier ces passages le terme *« étape »* sera utilisé dans la présentation des résultats des expériences. Dans l'exemple précédent, on distingue une *étape de requête initiale* d'une *étape de correction*.

5.1.3 Trois exemples de services

Les trois services présentés dans cette partie sont ceux qui ont été utilisés dans les expériences de la thèse : le premier (*'PlanResto'*, expérience 3) permet de trouver des restaurants à Paris ; le deuxième (*'Santiago'*, expériences 1 et 4) permet de prendre des rendez-vous avec les médecins d'un hôpital ; et le troisième (*'Cinéliste'*, expériences 2 et 5) permet de consulter une base de données de films. Pour les besoins expérimentaux de la thèse, ces services ont été réduits au nombre minimum de « phases » nécessaires pour assurer la réussite des « dialogues » dans le cadre expérimental. Par exemple, un service de prises de rendez-vous médicaux devrait pouvoir assurer un minimum de qualification des problèmes médicaux pour orienter vers le bon médecin. Mais le détail de cette problématique ne concerne pas l'approche générale de la thèse.

Dans les synoptiques des dialogues expérimentaux (Figures 5-2, 5-3 et 5-4, ci-dessous) les phases sont indiquées par des encadrements en lignes pointillées. Les actes qui composent ces phases sont représentés schématiquement sous forme de boîtes et flèches.

A Le service 'PlanResto'

Le dialogue avec le service *PlanResto* est présenté dans la Figure 5-2. Ce service permet de trouver un restaurant à Paris. Il existe en version web et en version téléphonique. Cette dernière a été développée à l'aide de la technologie *Artimis.* Elle démontre la souplesse de fonctionnement et l'aspect coopératif du dialogue dans un contexte où des conflits peuvent apparaître entre les critères de recherche et où de nombreuses réponses peuvent être proposées à l'utilisateur.

Le dialogue est organisé en trois phases :

- La *'phase de requête'* permet à l'utilisateur d'indiquer ses critères de recherche (au nombre de trois : *'type'*, *'lieu'* et *'prix'*). Si les critères apportés sont insuffisants, le système l'indique à l'utilisateur et l'invite à préciser sa demande ;

- Dans la *'phase de sélection'*, le système présente à l'utilisateur les réponses qui correspondent à sa demande. Il les présente une par une et invite chaque fois l'utilisateur, soit à demander plus d'informations, soit à passer à la réponse suivante, ou bien à recommencer avec une nouvelle recherche. Chaque réponse proposée indique le *'nom d'un restaurant'* et la *'gamme de prix'* des repas qui y sont servis. L'exploration totale de la liste dépend de l'initiative de l'utilisateur. Dans le service réel, des phases de désambiguïsation peuvent s'avérer nécessaires si trop de réponses sont disponibles. La version expérimentale proposait 6 à 10 réponses dans tous les scénarios ;

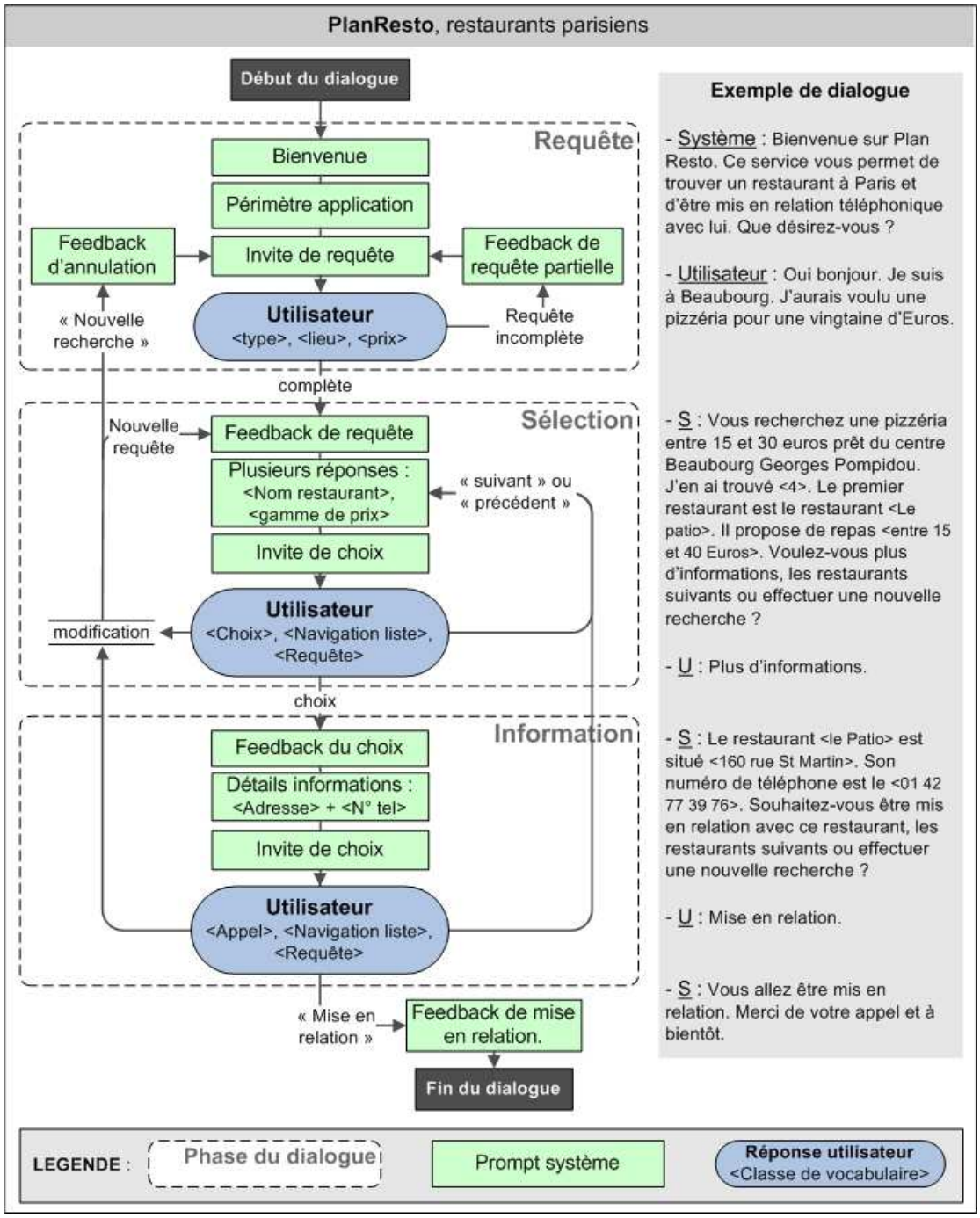


Figure 5-2 : Structure du dialogue du service *'PlanResto'*

- La *'phase d'information'* permet au système de présenter le détail des renseignements sur le restaurant choisit : *'nom du restaurant'*, *'prix'*, *'adresse'* et *'numéro de téléphone'*. Dans la version expérimentale, la demande de mise en relation clôturait le dialogue.

L'exemple de dialogue (à droite) précise le déroulement normal d'un dialogue avec ce service. Cependant, la structure du dialogue est ouverte et des liens entre les phases peuvent être activés à l'initiative de l'utilisateur.

B Le service 'Santiago'

Le dialogue avec le service *Santiago* est présenté dans la Figure 5-3. Ce service permet de prendre des rendez-vous avec les médecins d'un hôpital. Il n'a pas été développé dans sa forme industrielle (*i.e.* téléphonique) et n'existait que sous la forme d'une *spécification fonctionnelle*.

Le dialogue est organisé de la manière suivante :

- La *'phase de requête'* permet à l'utilisateur d'indiquer ses critères de recherche, au nombre de deux (*'nom du médecin'* et *'jour du rendez-vous souhaité'*). Dans un service complet, ces critères peuvent faire l'objet de négociation ; par exemple, si l'utilisateur ne connaît pas le nom du médecin qu'il doit rencontrer mais seulement sa spécialité. Pour les expériences, les critères de recherche étaient indiqués aux participants sur une fiche dédiée ;
- Dans la *'phase de choix'*, le système présente à l'utilisateur les réponses qui correspondent à sa demande et l'invite à sélectionner l'une d'entre elles. Ces réponses sont composées chacune d'une *date* et d'une *plage horaire*. La version expérimentale proposait 5 réponses dans tous les scénarios ;
- La *'phase de confirmation'* permet au système de rappeler la date et l'heure du rendez-vous choisit et de demander une confirmation à l'utilisateur. Si l'utilisateur confirme son choix, il clôture le dialogue. S'il l'infirme, le dialogue reprend au début de la liste des réponses, dans la phase de choix.

- ***Facteur de complexité : Erreur de reconnaissance vocale***

Pour manipuler la complexité des dialogues avec ce service, une ou deux erreurs de reconnaissance vocale ont été introduites dans certains dialogues (*Cf.* expériences 1 et 4). Ces erreurs étaient introduites de la manière suivante : après que le participant ait fini de compléter sa requête, le système proposait une liste de réponse pour un jour qui ne correspondait pas à la requête. Par exemple, si le participant disait : *« Je voudrais un rendez-vous avec le docteur Dubois le mardi 6 septembre. »* Le système répondait : *« Vous demandez un rendez-vous avec le docteur Dubois le jeudi 8 septembre… »* Le participant devait alors corriger l'erreur pour obtenir une liste de propositions correspondant à sa demande. Pour corriger, il pouvait formuler directement une nouvelle requête (*« correction directe »*) ou annuler avant de reformuler sa requête (*« annulation »*).

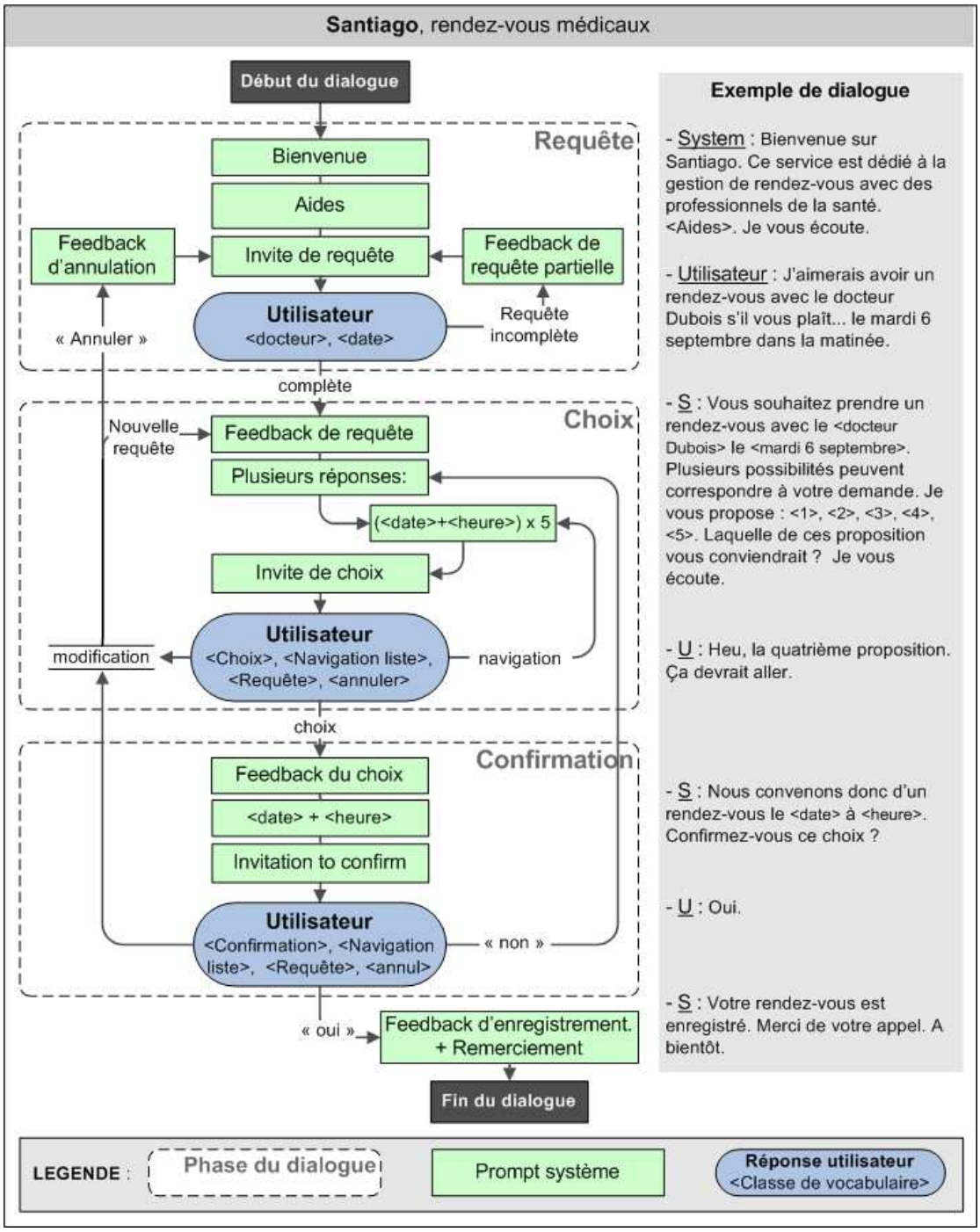


Figure 5-3 : Structure du dialogue du service *'Santiago'*

Comme pour le service précédent, un exemple de dialogue est présenté à droite du synoptique dans la Figure 5-3.

- ***Messages d'aide***

Les expériences réalisées avec ce service intégraient l'objectif de vérifier l'impact des messages d'aide. Le Tableau 5-1 présente les cinq messages d'aide utilisés. Au cours d'une cession expérimentale, un participant a entendu quatre de ces messages au maximum (voir expériences 1 et 4).

Ces messages d'aide (ou *prompts*) correspondent, dans leur style et dans leur contenu, à ceux qui sont utilisés dans des services en fonctionnement.

Tableau 5-1 : Aides du service Santiago

Type d'aide	Contenu	Durée *(vocale)*
'Barge in'	*« A tout moment, vous pouvez interrompre le système en lui coupant la parole. »*	**4** secondes
Formulation de requête	*« Vous pouvez faire votre demande de prise de rendez-vous soit de manière précise, par exemple : « Je souhaiterais prendre un rendez-vous le jeudi 8 septembre à 14H30. » Soit en restant évasif, par exemple : « Je souhaiterais prendre un rendez-vous en fin de semaine, plutôt en début d'après-midi. »*	**20** secondes
Correction d'erreur	*« Si je vous ai mal compris, vous pouvez répéter ou dire «annuler». Vous pouvez également directement corriger la partie d'un message que j'ai mal comprise. Par exemple, vous pouvez dire « Non, pas à 13h30 mais 16h30 »*	**16** secondes
Navigation dans la liste	*« Vous pouvez demander de réécouter la liste des propositions de rendez-vous, ou passer d'une proposition à la suivante ou à la précédente. Pour cela dites "début de liste" ou bien "la suivante", "la précédente" ou encore "réécouter la deuxième solution". »*	**15** secondes
Périmètre applicatif	*« Le dialogue se déroule en trois étapes. Vous faites d'abord votre demande en indiquant un professionnel de santé et une plage de disponibilité. Ensuite, le service vous propose les horaires disponibles et vous demande de choisir celui qui vous convient le mieux. Enfin, pour valider votre choix, vous devez le confirmer. »*	**19** secondes

C Le service 'Cinéliste'

Le dialogue avec le service *Cinéliste* est présenté dans la Figure 5-4. Ce service permet de consulter une base de données de films. Il n'a pas d'existence industrielle. Il a été imaginé et conçu pour répondre aux besoins expérimentaux de la thèse. Sa particularité est de fournir des données en plus grand nombre que les deux services précédents. Le dialogue est plus complexe, ce qui oblige les utilisateurs à parcourir les données et donne lieu à différentes stratégies d'exploration et à un test de rappel plus difficile. Le dialogue est organisé de la manière suivante :

- La *'phase de requête'* permet à l'utilisateur d'indiquer ses critères de recherche. De nombreux critères sont envisageables. Pour les expériences, seul le nom du réalisateur était utilisé. En effet, l'attention se portait sur la consultation des informations plutôt que sur la forme de la requête initiale ;
- Dans la *phase de choix*, le système présente à l'utilisateur la liste des films du réalisateur. Ces films sont classés en ordre alphabétique et numérotés. La version expérimentale proposait 10 films pour chaque réalisateur. Dans cette phase, l'utilisateur est invité à demander les détails du film qui l'intéresse. Il a également la possibilité de demander directement la lecture d'un film, s'il le désire ;

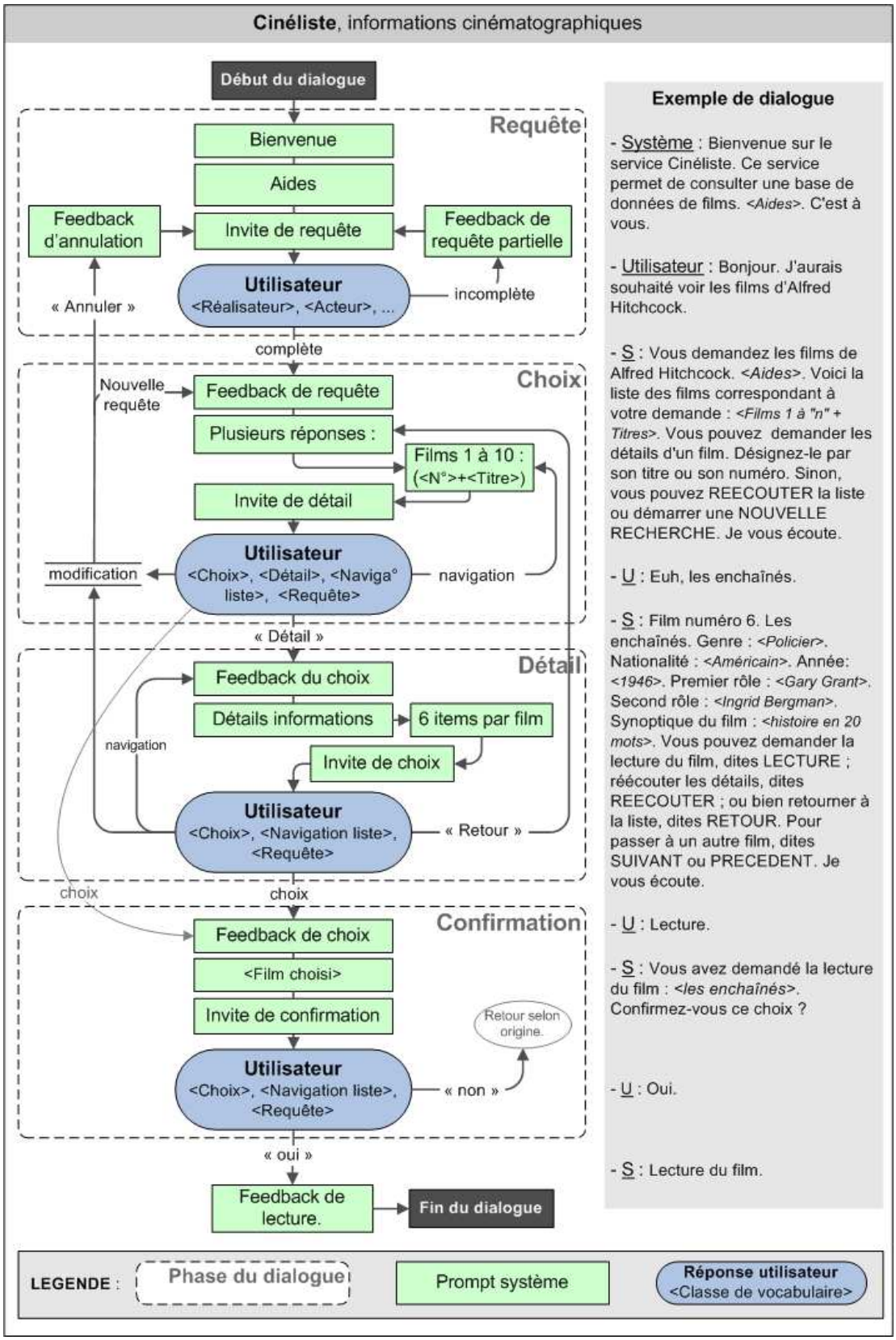


Figure 5-4 : Structure du dialogue du service *'Cinéliste'*

- Dans la *'phase de détail'*, le système présente à l'utilisateur un ensemble d'informations sur le film choisit : *'titre'*, *'genre'*, *'nationalité'*, *'année'*, *'premier rôle'*, *'second rôle'*, *'synopsis du film'*. Dans cette phase, il peut demander la lecture du film. Il peut également revenir à la liste des films (*'phase de choix'*) ou naviguer directement d'un film à l'autre

(au sein de la *'phase de détail'*) en utilisant les commandes *'suivant'* et *'précédent'* (ou leurs synonymes) ainsi qu'en déclinant le titre d'un film ou son numéro dans la liste ;

- La *'phase de confirmation'* permet au système de rappeler le titre du film choisit et de demander une confirmation. Si l'utilisateur confirme son choix, il clôture le dialogue. S'il l'infirme, le dialogue reprend soit dans la *'phase de choix'*, soit dans la *'phase de détail'*, selon l'historique du dialogue.

Un exemple de dialogue est présenté à droite du synoptique dans la Figure 5-4. Les hypothèses des expériences réalisées avec ce service ne portaient pas sur l'impact des messages d'aide. Cependant, le message d'accueil précisait brièvement le périmètre de l'application et présentait l'aide sur le *'barge in'* (*Cf.* Tableau 5-1) pour indiquer au participant qu'il pouvait se comporter naturellement.

- ***Facteur de complexité : Consigne***

Avec ce service, l'espace de recherche (contenu informationnel) est plus important que dans les deux services présentés précédemment. Pour cette raison, ce n'est pas l'erreur de reconnaissance vocale qui a été utilisée comme facteur de complexité, mais une consigne de recherche d'information :

- Dans le dialogue simple, le participant avait pour consigne d'identifier un film en particulier (*« Dans quel film Mme Baines meurt-elle ? »*) ;
- Dans le dialogue complexe, il devait trouver deux films (le plus ancien et le plus récent du réalisateur, parmi les films présentés), de sorte qu'il était obligé de se repérer plus profondément dans la liste des réponses pour en optimiser le balayage et y faire les comparaisons voulues.

5.1.4 Conclusion

La structure à deux niveaux hiérarchique qui a été utilisée pour décrire ces trois services permet de différencier plusieurs phases dans les dialogues (entre lesquelles un parcours est réalisé par tout utilisateur) et de décrire une structure semblable, dans chacune de ces phases. Cette structure comprend trois actes élémentaires – *écho, réponse et relance* – présents dans chacune des interventions du système.

La question qui se pose alors est de savoir si des manipulations de cette structure, au niveau inférieur, produisent des effets sur les parcours des utilisateurs, au niveau supérieur. On se demande également s'il en résulte des effets sur l'utilisateur lui-même, et notamment sur sa performance cognitive dans la tâche. L'enjeu est d'analyser les effets produits de façon à pouvoir en reproduire la mécanique.

5.2 Problématique

L'objectif de la thèse est double. Il consiste, d'une part, à étudier les déterminants du fonctionnement cognitif d'individus humains face à un système de DHM et, d'autre part, à étudier les règles de fonctionnement susceptibles de modifier le comportement interactif du système. La prise en compte de cette double contrainte a pour finalité d'identifier des moyens qui permettraient de réduire la complexité, pour l'utilisateur, liée au traitement des réponses du système. Pour cela, il est nécessaire d'assumer la complexité de l'activité dialogique, ce qui suppose, d'une part, de veiller à l'écologie des mises en situation expérimentales (voir Gibson, 1979; Rasmussen, Pejtersen & Goodstein, 1994), et d'autre part, de multiplier les indicateurs (voir, par exemple, Le Bigot et al., 2007) pour assurer un diagnostic aussi précis que possible.

Les formalisations de tâches présentées dans la partie précédente permettent de préciser les *degrés de liberté dont dispose le concepteur*. En effet, la complexité n'est pas une variable absolue, numérique, dont il suffirait de modifier la valeur. Les systèmes fonctionnent comme ils fonctionnent parce que les concepteurs font ce qu'ils savent faire. Réduire la complexité suppose d'identifier au préalable des techniques susceptibles d'opérer une réduction. Il ne s'agit pas là, en soi, d'un travail de recherche, mais d'un travail de conception, soit, de création. Or, la formalisation permet de représenter le fonctionnement actuel des systèmes et de réfléchir à des solutions alternatives. Plus particulièrement, les Figures 5-2, 5-3 et 5-4 ont permis de faire ressortir la similarité de structure des différentes *'phases'* qui composent les dialogues. Et c'est dans cette structure que des moyens ont été recherchés.

5.2.1 'Stratégie de présentation des informations' en 'DHM vocal'

Lors de la *conception d'un système de 'DHM vocal'*, la difficulté est de donner les moyens au système et/ou à l'utilisateur de gérer la double contrainte entre gestion du *processus interactif* et gestion des *connaissances* (voir Baker, 2004; Bunt, 1994). En ce qui concerne le processus interactif, la problématique de l'influence réciproque est très saillante (voir chapitre 3). A chaque instant, les choix comportementaux du système (*forme* des actes) contraignent l'utilisateur et lui offrent des opportunités d'action. La notion de *'stratégie'* est utilisée dans ce sens. Elle indique que la *'présentation des informations'* produit des effets divers dans l'interaction ; et elle suppose que ces effets peuvent être anticipés.

Conformément à la synthèse de la Figure 5-1, des effets peuvent être identifiés relativement à trois plans d'intentions (au moins) : communicatives (*e.g.* politesse, positionnement social, obligations de réponse), assertives (*e.g.* transmissions de connaissances), interactives (*e.g.* prises et cessions de parole, gestes, regards). Comme indiqué au chapitre 2, les modèles théoriques qui sont à la base du fonctionnement des systèmes de dialogue (modèles de type BDI) permettent de représenter les intentions du partenaire. Mais, par exemple, dans la

théorie de l'interaction (Sadek, 1991 ; Sadek, Bretier & Panaget, 1997, voir chapitre 2), seules les intentions *assertives* sont représentées. Le fonctionnement qui en découle est le suivant : le système *se fixe l'intention* de *détecter l'intention de l'utilisateur* d'obtenir une information, puis, ayant détecté cette intention, il *engage la procédure* qui permet de la satisfaire. Ce raisonnement est indépendant des dispositifs d'entrée et de sortie au travers desquelles l'intention est détectée ou satisfaite. Certaines *intentions communicatives* peuvent être intégrées au fonctionnement de certains systèmes (Keizer & Bunt, 2007, intègrent des règles de politesse). Mais les *intentions interactives*, portant sur des aspects de forme, ne sont formulées explicitement dans aucun système (à notre connaissance).

Pour cette raison, un travail d'analyse et d'expérimentation psychologique, permettant de d'évaluer parallèlement une variété d'effets internes au processus dialogique, et permettant de dissocier les fonctions premières qui causes ces effets, peut offrir de l'intérêt dans la perspective du développement des systèmes de dialogue. Ainsi, pour une *'stratégie de présentation des informations'* donnée, le problème est de savoir quels effets immédiats elle produit sur l'utilisateur et quelle performance globale en découle.

A Stratégie de présentation en mode auditif

Les SVI sont conçus pour le téléphone (*Cf.* chapitre 2). Dans ce cas, seul le mode auditif est disponible pour présenter l'information. Dans ces *services téléphoniques classiques* les éléments à présenter (*'écho'*, *'réponse'* et *'relance'*) sont structurés vocalement, c'est-à-dire temporellement, et sont donc présentés successivement. Ils ont une structure linéaire.

Lors de la conception de *'stratégies de présentation'*, les manipulations consistent à agir sur le *mode* de présentation, au sens large[1]. Dans le mode vocal, les manipulations consistent à modifier la formulation des prompts. Il est possible, soit de *supprimer une partie des informations* (les *aides*, dans l'expérience 1), soit de *modifier la formulation et le rythme* (expérience 2), de façon à simplifier la présentation. Les *« informations cible »*, dont on test le rappel en contexte expérimental, sont présentées dans tous les cas.

B Vers une stratégie de présentation bimodale

Toutes les informations présentées par un système de dialogue sont verbales. Cela donne l'avantage de permettre un choix, pour chaque *unité de contenu* et sans modification du contenu, entre (1) une *présentation visuelle*, (2) une *présentation auditive* et (3) une *présentation redondante* (ou *bimodale* : visuelle + auditive).

Les expériences 3, 4 et 5 testent des *stratégies de présentation bimodales* conçues en répétant ce choix sur les différentes unités de contenus identifiées. Dans ces stratégies, les *unités de contenu* sont, selon la relation établie par Coutaz et Nigay (1994, propriétés CARE) :

[1] Searle (1969) utilise le terme « mode » dans son sens linguistique. Le terme renvoie à toutes les modifications susceptibles de changer la « force illocutoire » de l'énoncé (*Cf.* chapitre 1). Par exemple, le fait de parler plus ou moins fort est un changement de mode tout autant qu'un choix lexical.

(1) *concurrentes*, (2) *assignées*, (3) *redondantes* ou (4) *équivalentes*. Les principes de conception utilisés sont présentés relativement à chaque expérience. L'objectif général est d'identifier les combinaisons qui auraient un impact positif dans le dialogue.

5.2.2 Principe d'évaluation des 'stratégies de présentation'

La thèse propose des analyses basées sur une diversité d'indicateurs (Le Bigot et al., 2007) afin de tenter un diagnostic élargi. En effet, parmi les études en psychologie citées dans la partie théorique, beaucoup tendent à proposer des interprétations construites à partir d'indicateurs peu diversifiés. Par exemple, les théories de l'apprentissage multimédia présentées au chapitre 3 sont construites uniquement à partir d'indicateurs de l'apprentissage (mémorisation, acquisition de schéma, tests de transfert). De même, de nombreuses études en dialogue sont basées principalement sur la durée de la tâche et sur la transcription verbale des dialogues qui sert à calculer des indicateurs verbaux (taux de pronoms personnels, nombre de mots, etc.). Chaque domaine tend à privilégier les indicateurs qui lui semblent les plus directement pertinents et à délaisser d'autres indicateurs qui pourraient être pris en compte. Ainsi, l'élargissement de la gamme d'observables pris en compte simultanément est un enjeu majeur pour la psychologie. Pour cela, la description fine des interactions est nécessaire. Notamment, dans le dialogue, la prise en compte des *comportements de prise et cession de parole* (*Cf.* Sacks, Schegloff & Jefferson, 1974) est importante pour étudier la combinaison *mot-geste* (Goffman, 1974) des énoncés des partenaires. De même, la confrontation des scores de rappel, des temps de consultation et des évaluations subjectives de difficulté (charge subjective) est importante pour la compréhension des résultats.

En vertu de ces principes, l'objectif de la thèse est de proposer des analyses larges, dont les conclusions seraient applicables à des systèmes de DHM destinés au grand public. Il est donc préférable de s'appuyer sur les indices disponibles dans ce type de système, pour assurer la reproductibilité de ces analyses. Mais par ailleurs, le cadre expérimental permet de recueillir des données supplémentaires, en dehors du fonctionnement du système, destinées au diagnostic scientifique. Ces deux types de données sont parfois nommés indicateurs *« on-line »* et *« off-line »* (*e.g.* Jamet, 2004).

- ***Indicateurs « on-line »***

Les données disponibles dans les systèmes de DHM sont principalement issues :

- De la *reconnaissance vocale*, qui transcrit la parole de l'utilisateur en une chaîne de caractères. On se trouve ici à un niveau de granularité inférieur à celui de la paire adjacente question-réponse. La chaîne de caractères est utilisée pour une analyse sémantique, qui relève de l'interprétation des *intentions assertives*. Il est également possible d'en extraire des interprétations relevant du *plan interactif*, à partir, par exemple des temps de réponse, de la longueur des énoncés, la vitesse d'élocution, etc. Dans le *plan communicatif*, il est possible d'interpréter certains contenus sociaux en fonction de leur signification : par exemple, des remerciements ou l'expression d'agacement ;

- Du *gestionnaire du dialogue*, qui identifie les choix de l'utilisateur dans les différents états et permet d'en retracer le parcours. Ces choix relèvent également de l'*interprétation assertive*. Ils renvoient à la sémantique des actions de l'utilisateur lors de sa progression dans le dialogue, dont les différents parcours dépendent. (Cette sémantique est représentée dans le système, notamment sous forme de *« classes syntaxiques »*). Il est possible que certaines *stratégies de présentation* induisent certains parcours, ce qui indiquerait que des modifications de forme (les *stratégies de présentation*, relevant du *plan interactif*) produiraient des effets sémantiques (le choix de classes d'actions desquelles résultent des choix syntaxiques, relevant du *plan assertif*). Il est donc possible d'observer des effets de causalité croisée. Par ailleurs, dans le plan strictement *interactif*, certains patterns comportementaux individuels peuvent être identifiés (comme le montrent Oviatt et al., 2004, au sujet de la synchronisation parole-geste).

Ces éléments permettent d'évaluer l'efficacité du système en termes de performance. Ils permettent également de mettre en place des processus inférentiel (détection des intentions) et de construire une représentation de la situation (des *« états mentaux »* dans la théorie de l'interaction de Sadek (1991). *Cf.* Chapitre 2).

- ***Indicateurs « off-line »***

Les indicateurs « off-line » sont recueillis après la fin des dialogues. Ils sont spécifiques au contexte expérimental et sont inconnus dans le cadre d'un système en fonctionnement. Leur évaluation expérimentale est destinée à permettre l'estimation de ces inconnues en contexte de fonctionnement réel. Dans les expériences de la thèse, les indicateurs « off-line » pris en compte sont le rappel des informations consultées et la charge cognitive subjective (questionnaires *'Workload Profile'* et *'NASA-TLX'*).

5.2.3 Question théorique : quel modèle explicatif ?

L'objectif est de comprendre en quoi consiste la réduction de complexité (potentielle) liée aux stratégies de présentation proposées. Deux modèles s'opposent :

(1) L'un *quantitatif*, basé sur les *'informations'* qui sont communiquées aux utilisateurs. La littérature utilise beaucoup la notion de *charge cognitive* (ou ses parents : *'charge de travail'* et *'charge de travail mentale'*) sur un mode additif. C'est le cas dans certains travaux sur l'apprentissage (*e.g.* Tindall-Ford, Chandler & Sweller, 1997). Et dans le domaine des interfaces vocales, il est assez largement admis que les énoncés vocaux (en sortie du système) provoquent une surcharge informationnelle pour l'utilisateur (*e.g.* Le Bigot, Rouet et Jamet, 2007 ; Walker & Whittaker, 2004) ;

(2) L'autre *qualitatif*, basé sur l'*analyse fonctionnelle*, d'inspiration pragmatique. La pragmatique envisage des effets divers liés à un ensemble de fonctions (ce qui renvoie à la notion d'*utilité*) des actes qui constituent les énoncés. Dans ce domaine, les prédictions principales sont rapportées aux interactions entre les individus (ou

« agents », *Cf.* Sadek, Bretier & Panaget, 1997). Cette approche est développée notamment par Allwood (*e.g.* 1995).

L'approche de la thèse consiste à proposer des expériences construites pour avoir une validité écologique suffisante et à rapprocher les résultats du modèle qui en donne la meilleure interprétation.

- ***Hypothèse cognitive : le 'traitement des informations'***

La question du traitement des informations par l'utilisateur, et de sa surcharge, renvoie à l'idée que les énoncés du système contiennent une trop grande *quantité d'informations* et/ou que les canaux perceptifs ne permettent pas de *faire pénétrer* toutes ces informations. Si l'on suit ce point de vue, il semble raisonnable de se demander si la suppression, l'échelonnement ou la simplification d'une partie des *'informations'*, ou la répartition sur plusieurs canaux, réduirait la charge et, par là, permettrait un traitement plus facile des informations et un meilleur apprentissage.

L'*hypothèse cognitive*, ou *'hypothèse additive'* (Mayer, 2005b ; Sweller 2005 ; Tindall-Ford, Chandler & Sweller, 1997), suppose que les différentes stratégies de présentation des informations ont pour conséquence de modifier la charge cognitive et que cette modification a pour conséquence de modifier l'apprentissage et le processus interactif. Elle conduit à la prédiction expérimentale qu'une manipulation du *'nombre d'informations'* ou du *'nombre de canaux perceptifs'* modifie la performance (apprentissage et processus collaboratif). Selon les auteurs qui défendent ce point de vue (*e.g.* Mayer, 2005; Schnotz & Kürschner, 2007; Sweller, 2005), les stratégies de présentation utilisées doivent être en accord avec l'architecture cognitive de façon (1) à identifier les différents types de charge qui pèsent sur cette architecture, (2) à identifier les traitements des différents modules (*« memory stores »*) et (3) à indiquer comment les différentes tâches chargent ces modules.

Selon cette hypothèse, la *charge cognitive* est une notion suffisante pour prédire la *performance* ; et la conception d'un algorithme permettant le *'choix dynamique de la stratégie de présentation'* peut reposer sur une formalisation de cette notion. Par exemple, l'*effet de modalité* indique que, pour des informations *graphiques* et *verbales*, la modalité visuelle doit être associée aux premières et la modalité auditive aux secondes, pour que la mémoire de travail ne soit pas surchargée.

- ***Hypothèse pragmatique : les 'fonctions de l'énoncé'***

L'approche pragmatique indique que l'énoncé est une succession d'actes. Pour tester l'efficacité d'un énoncé, il est nécessaire d'envisager les différents types d'effets qu'il produit et d'évaluer leur utilité relative dans la communication. Ces effets sont, selon le schéma de synthèse (Figure 5-1) des *effets interactifs* (relations spatio-temporelles entre les actions des partenaires), des *effets assertifs* (compréhension et mémorisation des contenus) et des *effets communicatifs* (relations interindividuelles). Dans cette approche, l'efficacité d'une stratégie dépend de la combinaison de ces différents types d'effets.

Cette hypothèse peut être désignée sous le nom d'*hypothèse pragmatique'* ou *'hypothèse fonctionnelle'*. Elle suppose que les actes remplissent différentes *fonctions* dans l'interaction, *i.e.* ils ont des finalités spécifiques. C'est la reconnaissance de ces fonctions qui permet à tout observateur d'un acte d'en inférer le sens. Ainsi, le processus interprétatif ne reposerait pas uniquement sur la *quantité d'informations* que traite le système cognitif. Au contraire, dans ce schéma, c'est l'interprétation de l'utilisateur qui lui permet d'extraire, par construction, les contenus qui lui sont présentés dans l'énoncé. Autrement dit, *'il construit les informations'* (voir, notamment, Merleau-Ponty, 1945) à partir de l'acte d'énonciation. Dans ce cas, l'objectif lors de la conception d'un énoncé serait de créer les conditions nécessaires à la construction des informations par le partenaire. Il s'agit de créer des opportunités.

Ainsi, la conception d'un algorithme permettant le *'choix dynamique de la stratégie de présentation'* reposerait sur l'analyse pragmatique des différents types d'effets possibles. Selon l'*hypothèse pragmatique*, les effets de modalité devraient être analysés sur la base de l'ensemble des fonctions que remplissent les constituants de l'énoncé dans une communication orientée vers une diversité de buts.

5.3 Présentation des expériences

Cinq expériences sont présentées dans la thèse. Les deux premières portent sur des dialogues dans lesquels le système communique uniquement en mode vocal. Dans les trois expériences suivantes, il utilise également la modalité visuelle et communique selon différents modes audio-visuels. Trois de ces expériences (les expériences 1, 3 et 4) ont été réalisées dans le cadre de conventions de recherche entre France Télécom R&D et l'Université Rennes 2. L'expérience 3 a été réalisée la première (Contrat N°42350346. Juin 2004). L'expérience 4 a été réalisée ensuite, suivie de l'expérience 1, dans le cadre d'une autre convention (Contrat N°46131031. Décembre 2006). Les expériences 2 et 5 ont été réalisées par la suite sans être encadrées par une convention de recherche. L'ordre de présentation choisi pour la thèse suppose une évolution des questionnements allant dans le sens de la problématique plutôt qu'une évolution chronologique.

Les cinq expériences portent sur les thèmes suivants dans le DHM :

- Mode vocal : (1) Effets des *aides procédurales* ;
  (2) Effets de la *verbosité du système* ;
- Mode bimodal : (3) Effets de la *redondance audio-visuelle* et *'effet de suffixe'* ;
  (4) Mise en évidence de la *spécificité modale* ;
  (5) *Sous-spécificité modale* et niveau d'analyse requis.

Le Tableau 5-2 donne un aperçu général des cinq expériences présentées. Le dispositif expérimental est présenté dans cette partie car il est identique dans toutes ces expériences.

Tableau 5-2 : Présentation synthétique des cinq expériences

N°	Objectif	Facteurs	Participants	Matériel
1	Rôle des aides pour le dialogue ? Pour quel coût ?	1 : Sans aide / Aides groupées / Aides réparties 2 : 'Complexité' *(erreurs de reconnaissance vocale)* 3 : 'Ordre' *(erreurs à l'essai 1 ou 2)*	69 (3 groupes)	*'Santiago'*
2	Rôle de la syntaxe ? Peut-on raccourcir les énoncés ? Impact et coût ?	1 : Syntaxe complète / réduite / réduite+aérée 2 : 'Complexité' *(consigne simple ou complexe)*	45 (3 groupes)	*'Cinéliste'*
3	Redondance audio-visuelle. Impact et coût ?	1 : *'Réponse'* Auditive / Visuelle / Redondante 2 : *'Relance'* A / V / R	54 (3 groupes)	*'PlanResto'*
4	Mise en évidence de la spécificité des modes de présentation.	1 : Stratégie *mono / bi modale* 2 : 'Complexité' *(erreur de reconnaissance vocale)*	80 (4 groupes)	*'Santiago'*
5	Quel niveau d'analyse de la spécificité modale ?	1 : Stratégie *bi / multi modale* 2 : 'Complexité' *(consigne simple ou complexe)*	48 (3 groupes)	*'Cinéliste'*

Chacune de ces expériences consistait à comparer plusieurs *stratégies de présentation*. La diversité des indicateurs pris en compte (voir les remarques proposées p. 138) permettait d'interpréter les résultats soit en référence aux théories de l'apprentissage (interprétation additive classique en psychologie cognitive), soit en référence aux théories pragmatiques (interprétation fonctionnelle issue de la pragmatique linguistique). Ainsi, l'intérêt de ces expériences était non seulement d'obtenir un diagnostic sur les *stratégies de présentation* étudiées, mais également de confronter les points de vue lors de l'interprétation pour enrichir le diagnostic et valider l'intérêt des différentes constructions théoriques pour alimenter ce diagnostic à un niveau conceptuel.

5.3.1 Dispositif expérimental

Les mises en situation expérimentales des cinq expériences de la thèse se voulaient écologiques. L'objectif n'était pas de reproduire totalement les conditions d'un dialogue avec un service en fonctionnement, mais de les reproduire d'une façon suffisante pour que les effets attendus puissent être observés. La technique utilisée dans les cinq expériences est le *Magicien d'Oz* (voir Fraser & Gilbert, 1991) car cette technique a été conçue dans ce but.

Pour l'expérimentation, il aurait été idéal de développer un système réel fonctionnant sur la base des formalismes proposés. Mais ce type de développement est coûteux (financièrement, temporellement, humainement et techniquement) et suppose de coordonner une équipe capable de développer les différents composants (reconnaissance vocale, gestion du dialogue, plate-forme téléphonique). Le *Magicien d'Oz* permet, à peu de frais (financiers, temporels, etc.), de reproduire le comportement d'un système de dialogue d'une façon

suffisante pour éprouver des hypothèses. Elle permet de prouver l'utilité des techniques proposées (*« proof of concept »*).

A Le protocole du 'Magicien d'Oz'

Le protocole en Magicien d'Oz (Dahlbäck, Jonsson & Ahrenberg, 1993; Fraser & Gilbert, 1991) permet de reproduire le fonctionnement d'un système de dialogue sans en implémenter toutes les composantes techniques. Il repose sur la programmation des réponses du système. Mais plutôt que de concevoir un modèle de reconnaissance vocale, c'est un humain (nommé *« Magicien d'Oz »*) qui écoute les commandes du participant/utilisateur et qui sélectionne les réponses à envoyer. De ce fait, il est nécessaire de limiter le nombre de réponses possibles pour faciliter la sélection par le *Magicien d'Oz*. En effet, au-delà d'une douzaine de possibilités, le *Magicien* doit contrôler trop de paramètres pour assurer une réponse rapide. Ainsi, les programmes de test ont été conçus pour reproduire le fonctionnement des services uniquement dans le cadre des scénarios expérimentaux. Les comportements du système étaient limités aux réponses nécessaires. En cas d'incompréhension ou de demande inappropriée de la part de l'utilisateur (demande hors périmètre de l'application), des messages d'erreur étaient prévus pour indiquer aux participants de reformuler leur demande.

B Dispositif technique

La Figure 5-5 présente le dispositif expérimental utilisé dans les cinq expériences de la thèse. Deux ordinateurs étaient nécessaires :

- ***PC 1*** (en bas à droite) : Le *Magicien d'Oz* disposait d'un clavier qui lui permettait de contrôler l'exécution du programme de test, sur l'ordinateur principal. Cet ordinateur permettait de simuler le système de DHM. (Le fonctionnement des programmes de test est décrit dans la section suivante). L'utilisateur écoutait les énoncés vocaux grâce à un micro-casque (indiqué par un téléphone dans la Figure 5-5). Il visualisait les énoncés présentés à l'écrit sur un écran (17 pouces). Un affichage jumeau de cet écran était également présenté au Magicien pour qu'il dispose des moyens de contrôle nécessaires (non représenté dans la Figure 5-5). Les énoncés vocaux du système étaient des messages préenregistrés avec une voix de synthèse (synthèse vocale *France Télécom ;* Agnès® v. 4.3; 16 kHz). ;
- ***PC 2*** (au centre) : Les réponses du participant (utilisateur), produites vocalement dans le micro-casque, étaient enregistrées de façon synchrone avec les énoncés du système sur le second ordinateur. L'enregistrement sonore était présenté au *Magicien d'Oz* (casque, non représenté dans la Figure 5-5), ce qui lui permettait d'écouter ces réponses pour interpréter la commande correspondante et de sélectionner la touche clavier correspondante pour recommencer le cycle.

L'utilisation de ce dispositif supposait une phase d'entraînement préalable du *Magicien d'Oz* pour assurer des réponses à la fois précises et rapide.

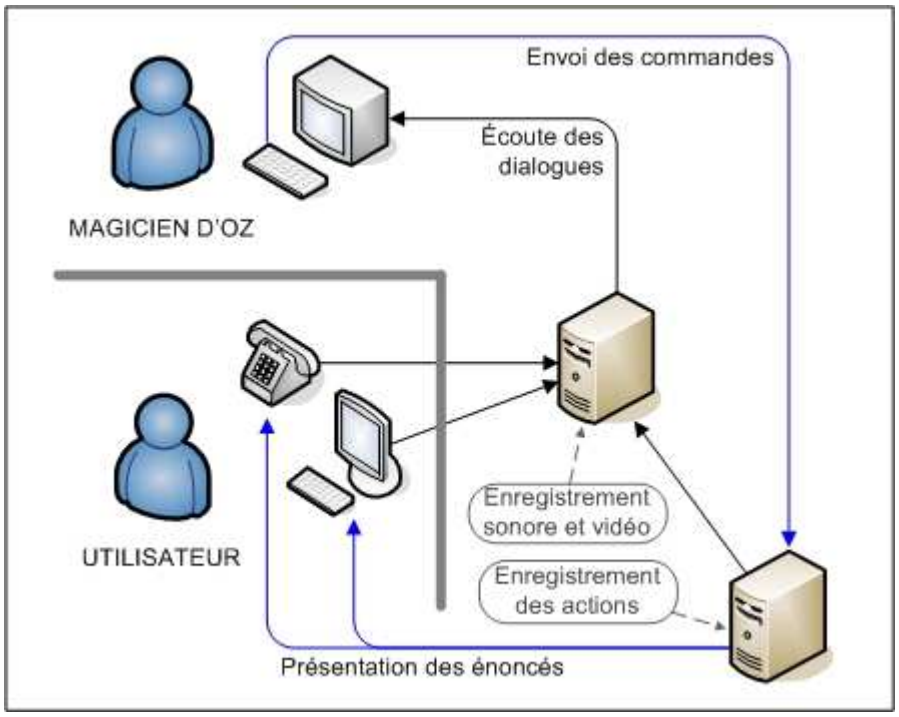


Figure 5-5 : Dispositif technique utilisé pour le protocole en 'Magicien d'Oz'

C Développement des programmes de test

Le développement des programmes de test des cinq expériences a été réalisé avec Macromedia Director MX®. Ils étaient exécutés sur un ordinateur équipé de Windows XP® et doté d'un processeur de 1,7 GHz et de 256 Mo de mémoire vive.

Macromedia Director permet de programmer des scénarios qui exécutent des actions (scripts). Il permet également d'associer des entrées du système (clavier ou souris) pour contrôler ces scripts. Pour les expériences, les touches du clavier étaient utilisées car, après apprentissage, elles permettent une sélection plus rapide. Chaque réponse du système était associée à une touche du clavier. Par exemple, dans la copie d'écran présentée dans la Figure 5-6, les films d'Alfred Hitchcock étaient présentés à l'utilisateur. S'il demandait les DETAILS du film *« Fenêtre sur cour »*, le Magicien d'Oz sélectionnait la touche *« 1 »* ; LIRE ce film, c'était la touche *« a »* ; faire une NOUVELLE RECHERCHE, touche *« w »*, etc.

Les programmes étaient conçus pour qu'une seule touche du clavier permette d'envoyer la réponse appropriée. Dans chaque phase, les réponses inappropriées étaient désactivées de façon à éviter les erreurs. Pour une réponse rapide, le Magicien d'Oz devait se repérer correctement dans le dialogue. En cas de doute, il disposait d'une fiche récapitulant les touches correspondantes aux actions disponibles dans les différents états. Après entraînement, toutes les réponses étaient sélectionnées en moins de deux secondes.

- ***Présentation visuelle des informations***

La Figure 5-6 présente une copie d'écran du téléphone utilisé pour présenter les informations aux participants des expériences.

Figure 5-6 : Téléphone simulé sur PC

La présentation visuelle des données (cas des expériences 4, 5 et 6) était structurée en trois zones :

(1) La *'partie supérieure'* correspondait aux ***'échos'***. Dans cet exemple issu de *Cinéliste* le nom du réalisateur demandé était affiché ;

(2) La *'partie centrale'* était l'espace des ***'réponses'***. Dans l'exemple, c'était la liste des films du réalisateur ;

(3) La *'partie inférieure'* présentait les ***'relances'*** disponibles, c'est-à-dire les commandes que l'utilisateur pouvait utiliser.

NB : Dans les différentes stratégies testées, une partie seulement de ces informations était affichée de façon à en éprouver l'utilité et les effets.

D Spécificités liées au contexte expérimental

- ***« Reconnaissance vocale »***

Aucun programme de reconnaissance vocale n'était utilisé dans ce dispositif expérimental. Le *Magicien d'Oz* agissait virtuellement comme un automate de reconnaissance vocale.

D'un coté, un *Magicien d'Oz* « fonctionne » mieux qu'un automate puisqu'il est capable d'une meilleure compréhension de la parole de l'utilisateur. Ainsi, aucune erreur de reconnaissance vocale n'apparaissait dans ce contexte (excepté celles qui ont été introduites par manipulation expérimentale). Tout vocabulaire compréhensible dans le cadre de la tâche était interprétable grâce à cette oreille humaine. Ainsi, dans ce contexte expérimental, la reconnaissance vocale avait un « fonctionnement parfait ». Cela apporte une limite en ce qui concerne l'approche écologique ; et la généralisation des résultats suppose la mise en place de tests dans des systèmes réels.

D'un autre coté, un *Magicien d'Oz* est un humain et il peut produire d'autres types d'erreurs. Par exemple, il est capable de faire preuve d'une empathie dont n'est pas capable un système, et d'utiliser ses connaissances pour produire les actions attendues plutôt que les actions commandées par l'utilisateur. Pour cette raison, la phase d'apprentissage avait également pour but d'apprendre à ne répondre qu'aux commandes, sans les anticiper. En pratique, cette remarque est très importante. Au cours de la thèse, un étudiant mis en position de *Magicien d'Oz* et insuffisamment briefé a fait preuve de ce type d'empathie envers

les participants, de sorte que la totalité des résultats a due être mise de coté et que la totalité des passations a due être renouvelée.

- ***Comportement en cas d'erreur***

Lorsque la demande du participant n'était pas compréhensible dans le cadre de la tâche, le Magicien d'Oz sélectionnait l'énoncé : *« Je n'ai pas compris. Veuillez reformuler votre demande s'il vous plaît. »* Dans un système en fonctionnement, des comportements plus riches peuvent être associés à des erreurs de différents types qu'il est difficile de différencier en dehors d'un automate de reconnaissance vocale. Dans ce cadre expérimental, l'objectif n'était pas de traiter ces cas, mais de les éviter. Lorsque ce type de « déviance » n'a pu être évité et s'est répété dans le dialogue, les résultats correspondants n'ont pas été conservés.

- ***Calibrage des temps***

Certaines remarques spécifiques peuvent être apportées sur le calibrage des énoncés dans les différentes stratégies de présentation. Elles sont précisées relativement à chaque expérience, lors de la présentation de la méthode, dans la partie dédiée au protocole expérimental, qui précise les conditions expérimentales testées.

5.3.2 Hypothèse générale

Les manipulations expérimentales qui ont été construites permettent soit une interprétation cognitive, en termes de traitement des contenus présentées, soit une interprétation pragmatique, en termes fonctionnels liés aux effets des énoncés :

- Selon le *point de vue cognitif*, si moins d'information est présentée, ou si plus de capacités perceptives (plusieurs canaux sensoriels) peuvent être investies, des effets peuvent apparaître du fait d'une plus forte disponibilité d'énergie attentionnelle (on parle de *'métaphore énergétique', e.g.* Barrouillet, 1996). Dans cette perspective, la manipulation des stratégies de présentation du système doit faire baisser la charge cognitive de l'utilisateur et cette baisse doit libérer des ressources qui permettent une meilleure mémorisation des informations et un meilleur comportement adaptatif.
- Selon le *point de vue pragmatique*, si des modifications sont apportées dans les stratégies de présentation, elles réalisent des fonctions dialogiques différentes qui peuvent avoir des conséquences différentes pour le processus collaboratif. Ainsi, ces stratégies modifieraient l'adaptation réciproque entre les partenaires en communication, ce qui pourrait produire des effets sur les processus interprétatifs mis en place (charge cognitive, mémorisation, utilité perçue) par chacun des partenaires.

L'hypothèse est que l'*analyse pragmatique* fournit une interprétation plus riche et peut permettre de dégager un *modèle explicatif* utilisable pour prédire la performance, soit les effets des actes.

Chapitre 6 Dialogue homme-machine vocal

Si les systèmes de DHM surchargent l'utilisateur lorsqu'ils lui présentent l'information sur la modalité auditive (Le Bigot et al., 2007), il semble raisonnable de chercher des solutions pour faire baisser la charge cognitive en agissant sur la quantité d'information présentée.

L'*expérience 1* a pour but de vérifier si la suppression des messages d'aide peut permettre d'abaisser la complexité des traitements imposés aux utilisateurs des systèmes de DHM. On se demande quel coût est associé au traitement de ces messages, mais aussi quels autres effets apparaissent. La question d'arrière plan est de savoir si ces messages sont utiles et/ou nécessaires dans un système de DHM.

L'*expérience 2* a pour but de vérifier si la suppression d'une partie de la syntaxe dans les énoncés du système peut permettre des gains d'efficacité. On se demande quel coût est associé à la syntaxe des messages du système et à sa suppression. La question d'arrière plan est de savoir si une syntaxe rigoureuse est nécessaire ou s'il est possible de faire composer une syntaxe moins performante à un système de DHM sans gêne pour l'utilisateur.

6.1 Expérience 1 : Effet des messages d'aide

6.1.1 Objectifs et hypothèses

Les messages d'aide sont des messages déclaratifs et/ou procéduraux qui permettent de présenter à l'utilisateur des informations d'ordre général sur le service lui-même (périmètre applicatif), ou plus particulières, sur son utilisation (commandes disponibles, libertés et contraintes comportementales, limites du système).

Dans les systèmes de DHM en fonctionnement, il existe plusieurs modes de présentation de ces messages d'aide. (1) L'une des solutions consiste à les présenter en début de dialogue, dès le message d'accueil. (2) Une autre solution consiste à indiquer à l'utilisateur que ces informations sont disponibles sur demande (touche « * » dans des services DTMF ou en prononçant le mot « aide » dans des services en reconnaissance vocale). (3) Enfin, il est possible de distribuer l'information et de ne la présenter que lorsqu'elle est utile de façon à ce que l'utilisateur dispose toujours des informations dont il a besoin au bon moment. Cette solution peut sembler la plus prometteuse, mais Bubb-Lewis et Scerbo (2002) ont aussi montré que cela peut être coûteux du point de vue du contrôle du dialogue par l'utilisateur et du temps consommé. De plus, cette solution est également la plus contraignante du point de vue de la conception des services (spécification initiale et contraintes de programmation).

Les messages d'aide testés dans cette expérience ont un contenu de nature procédurale, portant sur le déroulement du dialogue. Du point de vue théorique, on peut se demander quels traitements cognitifs sont associés à ces messages. L'approche cognitive envisage ces traitements comme un coût dont la dépense est assurée par des pools de ressources. La théorie de la charge cognitive envisage les différentes sources de coût sur un *mode additif*. Selon ce point de vue, les ressources cognitives existent en quantité limitée et des ressources peuvent être libérées tant que toutes ne sont pas consommées. L'approche pragmatique tend au contraire à envisager les traitements cognitifs sur un *mode fonctionnel.* Dans ce cas, les messages du système sont vus comme des indices qui permettent à l'utilisateur de construire la signification de la situation pour s'orienter dans l'action. Le coût des messages peut alors être vu comme une sous-partie des effets qu'il produit ; et l'objectif est d'évaluer ce coût et ces effets aussi largement que possible. En effet, des auteurs tels que Sheeder & Balogh (2003) ont pu montrer l'impact des messages d'aide sur la requête formulée immédiatement après le message d'accueil. Ici, l'intérêt est d'étudier les effets produits dans l'ensemble du dialogue et sur une diversité d'indicateurs.

Les aides sur demande (solution (2) ci-dessus) sont généralement peu utilisées dans les services. La mise en place de cette condition en contexte expérimental est problématique car elle ne donne lieu qu'à un faible nombre d'observations (voir, par exemple, Babin, 2007). Ici, cette condition n'est pas testée car l'objectif n'était pas de savoir si les aides sont utilisées,

mais d'en étudier les effets. Cette expérience a été conçue pour acquérir des repères sur les bénéfices qu'il est possible de rechercher lors de la conception de messages d'aide. L'apport d'informations supplémentaires telle que celles que propose une aide impose-t-il une surcharge cognitive à l'utilisateur ? Cette charge gêne-t-elle la mémorisation des contenus ? En quoi consistent les bénéfices et les coûts des aides ? Leur distribution dans le dialogue est-elle nécessaire ?

- ***Point de vue cognitif***

Suivant le *point de vue cognitif*, les énoncés du système contiennent des informations dont le traitement impose une charge cognitive aux utilisateurs. Les messages d'aide génèrent une *charge cognitive* supplémentaire, qui peut perturber la mémorisation des informations cible. Ainsi, d'après la théorie de la charge cognitive :

(1) Les *informations procédurales* contenues dans les messages d'aide sont liées à l'activité d'*atteinte du but*. Elles génèrent une charge cognitive qui n'est pas directement liée à l'activité d'*acquisition de schémas*. De ce fait, quand des aides sont présentées, moins de capacité cognitive est disponible pour l'acquisition de schémas et celle-ci pourrait être perturbée. Selon ce principe, les informations cible (qui correspondent aux schémas à acquérir) pourraient être moins bien mémorisées quand des aides sont présentes.

(2) Par ailleurs, certains travaux indiquent que la *charge cognitive inutile* peut être réduite si les informations sont présentées pas-à-pas (*e.g.* Clarke, Ayres & Sweller, 2005; Kester, Lehnen, Van Gerven, & Kirschner, 2006). L'acquisition séquentielle (une à une) des informations devrait permettre à l'utilisateur de réduire sa *charge cognitive inutile* – liée à une absence de contiguïté temporelle qui crée un effet de partage de l'attention – et de la convertir en une *charge cognitive utile* – qui facilite la mise en relation des informations. Ainsi, la théorie de la charge cognitive indique que la répartition des aides à différents moments du dialogue devrait permettre une meilleure compréhension de leur contenu.

(3) Enfin, les erreurs de reconnaissance vocale induisent la répétition de certaines phases du dialogue, ce qui génère également une charge cognitive inutile et pourrait faire baisser la performance de rappel des informations.

- ***Point de vue pragmatique***

Suivant le *point de vue pragmatique*, les aides sont des actes supplémentaires au sein des énoncés du système. Elles peuvent provoquer divers effets dans la communication.

(1) Concernant l'apprentissage des contenus, aucune concurrence n'est supposée *a priori* entre le traitement des contenus procéduraux présentés dans les aides et des autres contenus présentés dans les prompts. Au contraire, les aides sont susceptibles d'apporter aux participants des indications qui leur permettent de construire une représentation exacte du système afin d'assurer un meilleur contrôle du dialogue et ainsi de focaliser leur attention sur les contenus recherchés. Selon ce point de vue, la présence des aides ne doit pas gêner la mémorisation des informations.

(2) La présentation pas-à-pas correspond au *modèle de contribution* (Clark & Wilkes-Gibbs, 1986 ; Clark & Schaefer, 1989) dans lequel l'accord sur les références fait l'objet de propositions et de validations successives. Selon ce modèle, la distribution des aides dans les phases du dialogue correspondantes permet de prendre en compte immédiatement leur contenu, comme c'est le cas dans les phases de *présentation* et d'*acceptation* du modèle.

(3) En cas d'erreur de reconnaissance de la part du système, les participants qui ont entendu les aides devraient s'adapter plus facilement et être plus efficace. Pour les participants qui n'ont pas entendu les aides, il est possible que leurs réactions (inférences et réponses comportementales) aient un effet négatif sur le rappel et sur les évaluations subjectives concernant le système.

Les prédictions émanant de ces deux perspectives ne sont pas nécessairement contradictoires. Elles apportent un point de vue cognitif, focalisé sur les processus liés à l'apprentissage, et un point de vue sociocognitif, qui privilégie le processus interactif en jeu dans la communication.

6.1.2 Méthode

A Participants

Le nombre final de participants était de 69 (17 hommes, 52 femmes). La moyenne d'âge du groupe était de 21,91 ans (*é.t.* = 2,92 ; *min.* = 18,33 ; *max.* = 33,5). Les participants étaient des étudiants l'Université Rennes 2 en provenance de plusieurs disciplines et s'étalant de la première année d'étude après le bac au Master 2 (*moyenne* = 2,54 ; *é.t.* = 1,12). Chaque participant a reçu un bon d'achat d'une valeur de 10 € pour sa participation.

Le niveau d'expérience en informatique des participants était moyen. Ils ont déclaré passer en moyenne 13,23 heures par mois à utiliser un ordinateur (*é.t.* = 15,03 ; *min.* = 0 ; *max.* = 80) dont 7,07 sur internet (*é.t.* = 11,22 ; *min.* = 0 ; *max.* = 80). Pour comparaison, ils ont déclaré passer 10,20 heures par mois à regarder la télévision (*é.t.* = 7,49 ; *min.* = 0 ; *max.* = 35). En ce qui concerne la pratique des services téléphoniques, leur niveau d'expérience était peu élevé. Il portait principalement sur l'utilisation de leur répondeur téléphonique. Certains participants ont déclaré avoir déjà utilisé des services grands public, tels que le service des renseignements (question oui/non).

B Matériel

Le système utilisé pour cette expérience était une simulation du service *'Santiago'* (*Cf.* Figure 5-3) développée et mise en place selon le dispositif expérimental présenté au chapitre 5. Ce service correspond à un dialogue en trois phases (*'Requête' – 'Choix' – 'Confirmation'*) dédiées à la construction d'une référence commune : un rendez-vous. Cette structure du dialogue fait de ce service un prototype idéal pour l'expérimentation car ces trois phases sont

nécessaires à tout dialogue de recherche d'information. La structure des échanges dans ce service est représentative de nombreux services vocaux.

C Procédure

Le participant était accueilli individuellement. Il était conduit dans une pièce calme où était installé le matériel expérimental. L'expérimentateur lui présentait brièvement, à l'oral, le service *'Santiago'* et le matériel qu'il s'apprêtait à utiliser pour réaliser les appels. Il lui remettait ensuite deux cahiers (*Cf.* Annexes, p. 259) :

(1) L'un pour les *consignes liées à l'utilisation du service*. Il incluait quatre pages : une page de présentation générale, et trois pages de consignes pour les trois appels à réaliser auprès du service. Chacune indiquait le *nom du médecin* avec lequel un rendez-vous était souhaité et présentait un *planning* (du mercredi 31 août au mercredi 7 septembre), indiquant les *plages de disponibilité* parmi lesquelles le rendez-vous devait être placé.

(2) L'autre pour les *consignes liées à l'évaluation de la charge cognitive*. Il incluait une page de présentation du questionnaire *'Workload Profile'* et une page de présentation de chacune des dimensions à évaluer. Ce cahier n'a été remis qu'aux participants qui ont évalué leur charge cognitive avec le questionnaire *'Workload Profile'*. Le questionnaire *'NASA-TLX'* n'inclut pas ce type de description des dimensions à évaluer.

Après avoir pris connaissance de ces documents, le participant commençait la procédure expérimentale sur l'ordinateur dédié. La procédure était organisée par le programme de test qui présentait le questionnaire d'évaluation de la charge subjective suite à chaque dialogue et un questionnaire final pour demander au participant d'évaluer l'utilisabilité du système et pour obtenir les renseignements généraux le concernant.

Après avoir terminé cette procédure, une feuille de papier était fournie au participant, présentant un tableau à neuf cellules (colonnes : jour, date, heure ; lignes : appel 1, appel 2, appel 3) précédé de la question : *« Vous avez pris trois rendez-vous avec trois médecins. Quels sont le jour, la date et l'heure de ces rendez-vous ? »* Aucune consigne de rappel n'avait été donnée préalablement.

D Protocole expérimental

Trois variables indépendantes ont été manipulées dans cette expérience.

- ***Facteur 1*: Condition d'aide**

Trois conditions sont évaluées, correspondant à trois « stratégies d'aide » :

1. ***'Sans aide'*** : Aucun message d'aide n'est présenté à l'utilisateur. Le message d'accueil indiquait seulement que *« le service est dédié à la gestion de rendez-vous avec des professionnels de la santé »* (voir la transcription complète dans la Figure 5-3). Il a une durée de 12 secondes.

2. ***'Aides groupées'*** : Quatre messages d'aides étaient présentés dans le message d'accueil : (1) périmètre applicatif, (2) 'barge in', (3) correction d'erreur et (4) navigation dans la liste. L'ajout de ces prompts (total : 54 secondes) portait la durée du message d'accueil à 68 secondes.

3. ***'Aides réparties'*** : Deux messages d'aide étaient présentés dans le message d'accueil : (1) périmètre applicatif et (2) 'barge in'. Cela portait la durée du message d'accueil à 35 secondes. Les deux autres messages étaient présentés immédiatement après l'écho de la requête de l'utilisateur, lorsque celle-ci était complète : (3) correction d'erreur et (4) navigation dans la liste. De ce fait, la durée de 11 secondes nécessaires pour présenter l'écho était prolongée de 31 secondes (soit 42 secondes au total) avant de présenter la liste des réponses au participant.

Il s'agit d'une *variable intergroupe*. Trois groupes ont été constitués, correspondant aux trois conditions. Ainsi, chaque participant a rencontré une seule condition au cours de l'expérience.

- ***Facteur 2 : Essai avec ou sans erreurs de reconnaissance vocale***

Chaque participant a réalisé trois dialogues expérimentaux. Le troisième dialogue incluait une condition supplémentaire de désambigüisation de la requête, liée à la spécialité médicale du médecin demandé. Il n'est pas inclut aux résultats car cette condition complique le protocole et n'apporte pas d'informations supplémentaires. Deux essais sont considérés :

- ***'Essai sans erreur'*** : L'un des essais s'est déroulé sans perturbation. Toutes les commandes/demandes du participant étaient correctement prises en compte ;
- ***'Essai avec erreurs'*** : Pour l'autre essai, ***deux erreurs*** *de reconnaissance vocale* étaient artificiellement introduites après la requête du participant.

Il s'agit d'une *variable intragroupe*. Chaque participant rencontrait les deux conditions au cours de deux essais successifs. La position du dialogue avec erreur était contre-balancée entre l'essai 1 et l'essai 2. (Le dialogue supplémentaire n'était pas inclut dans ce contre-balancement et se trouvait toujours en position 3).

- ***Facteur 3 : Position du dialogue avec erreurs***

La position du dialogue avec erreurs de reconnaissance vocale a été prise en compte comme variable indépendante :

- ***'Erreurs à l'essai 1'*** : Les erreurs ont été introduite dès le premier essai ;
- ***'Erreurs à l'essai 2'*** : Les erreurs ont été introduites au second essai.

Il s'agit d'une *variable intergroupe*. Dans chacune des conditions, la moitié des participants a subit ces erreurs à l'essai 1 et l'autre moitié les a subit à l'essai 2.

E Mesures dépendantes

Les mesures utilisées dans la présentation des résultats sont indiquées dans le Tableau 6-1. Ces indicateurs sont présentés plus en détail à la fin du chapitre 6.

Tableau 6-1 : Indicateurs utilisés pour la présentation des résultats de l'expérience 1

Rappel	Nombre d'items rappelés parmi 9 (*'jour de la semaine'*, *'date'* et *'heure'* des trois rendez-vous).
Charge cognitive subjective	Evaluée avec le questionnaire *'NASA-TLX'* pour une partie des sujets et *'Workload Profile'* pour l'autre. Dans le questionnaire *WP*, deux échelles liées au stress ont été ajoutées (issues de Lazarus & Folkman, 1984) : (1) *Sentiment de frustration*, (2) *Sentiment de perte de contrôle*.
Nombre de tours de parole (TDP)	Nombre total de prises de parole par le participant au cours du dialogue.
Nombre de mots	Nombre total de mots prononcés par le participant au cours du dialogue.
Durée	Durée totale du dialogue (en secondes).
Moment de la prise de parole	Catégorisée relativement à l'énoncé du système : (1) Dès l'écho, (2) Pendant la liste, (3) Après la relance.
Mode de correction de l'erreur	(1) Correction ou (2) Annulation
Nb de mots pour la correction	Nombre de mots pour corriger la première erreur.

6.1.3 Résultats

Le corpus final est composé de 138 dialogues au cours desquels 1145 commandes des participants ont été prises en compte pour un total de 8834 mots. Le nombre moyen de tours de parole (soit le nombre d'énoncés prononcés pour commander le système) était de 5,05 pour le dialogue sans erreur de reconnaissance vocale et de 11,54 pour le dialogue avec deux erreurs. Les analyses statistiques réalisées sur les données sont précisées relativement à chaque indicateur présenté.

A Indicateurs cognitifs généraux

La performance de rappel a été analysée avec une procédure GLM (*« General Linear Model »*) utilisant la *'condition d'aide'* (facteur 1) et la *'position de l'essai avec erreur'* (facteur 3) comme facteurs catégoriels. Le nombre de mots, le nombre de tours de parole et la durée des deux dialogues expérimentaux ont été inclus en co-variables.

La même procédure GLM a été appliquée aux indices globaux de charge cognitive. Une évaluation de charge étant disponible relativement à chaque dialogue, le facteur essai (facteur 2) a été ajouté et traité en intragroupe, en tant que mesure répétée.

- ***Rappel des rendez-vous pris au cours de l'expérience***

Comme on le voit sur le graphique de la Figure 6-1, le rappel des informations a été plus faible quand les participants ont subit les erreurs de reconnaissance vocale dès le premier essai. En l'absence d'aide, la position de l'essai avec erreurs a joué un rôle peu important. Une différence semble être apparue dans la condition *'aides groupées'* et elle a été très

importante dans la condition *'aide réparties'*. Cette dernière condition a donné lieu au meilleur rappel, mais elle a également été la plus sensible à la position du dialogue avec erreurs.

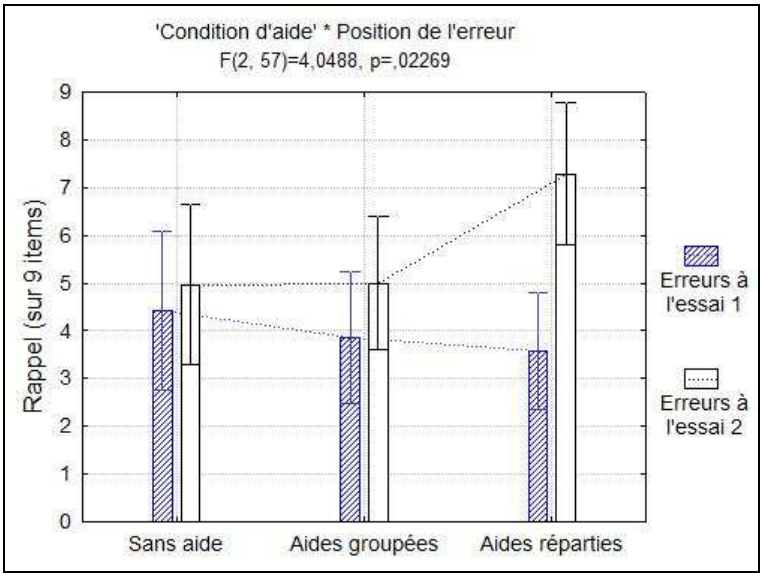


Figure 6-1 : Performance de rappel des informations

Le test de Levene révèle des distributions homogènes : $F(5,63)$ = 0,79 ; p=.55

La condition d'aide n'a pas eu d'effet sur le rappel des rendez-vous pris au cours des expériences :

$F(2,57)$=1,18 ; p=.31 ; η^2_p=.03

La position du dialogue avec erreurs a eu un effet :

$F(1,57)$=10,03 ; p<.01 ; η^2_p=.15

On constate également un effet d'interaction entre ces deux facteurs :

$F(5,57)$=4,04 ; p<.05 ; η^2_p=.12

- ***Indices de charge cognitive***

Les résultats obtenus avec TLX ont révélé un effet d'interaction significatif entre les facteurs 2 et 3 ($F(1,21)$ = 8,03 ; p<.01 ; η^2_p=.27). Quand les erreurs étaient à l'essai 1, leur disparition à l'essai 2 a fait baisser les évaluations subjectives de charge ; et quand il n'y avait pas d'erreur à l'essai 1, leur apparition à l'essai 2 a également fait baisser les évaluations subjectives de charge. Cet effet d'interaction indique que les évaluations de charge (avec TLX) ont été sensibles à un effet d'ordre (qui n'était pas considéré parmi les facteurs expérimentaux). Aucune autre différence n'est apparue dans les évaluations faites avec NASA-TLX. Ces résultats ne sont pas détaillés dans la mesure où les résultats obtenus avec *'Workload Profile'* ont eu une plus grande sensibilité aux facteurs expérimentaux.

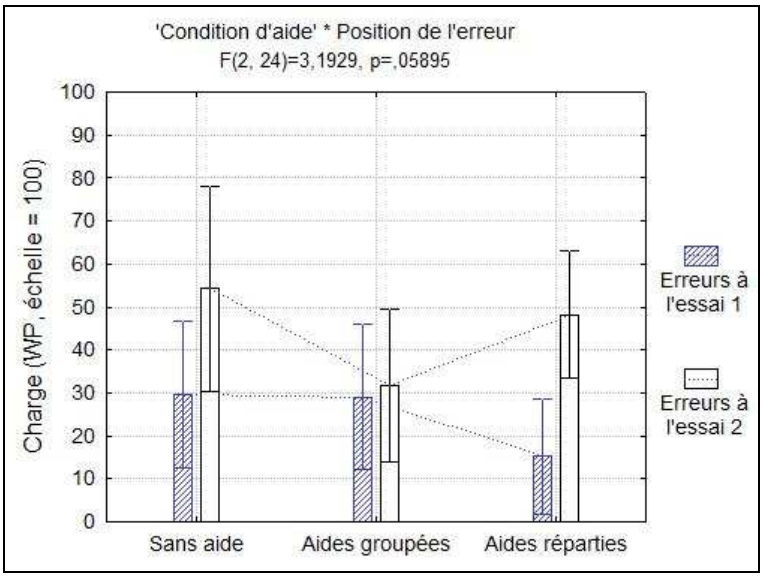


Figure 6-2 : Charge subjective pour l'essai avec erreurs

La condition d'aide n'a pas eu d'effet sur l'évaluation de charge :

$F(2,24)$=0,38 ; p=.68 ; η^2_p=.03

La présence des erreurs n'a pas non plus eu d'effet :

$F(1,24)$=2,74 ; p=.11 ; η^2_p=.10

En revanche, la position du dialogue avec erreurs a eu un effet :

$F(1,24)$=10,46 ; p<.01 ; η^2_p=.30

On constate également un effet d'interaction marginal entre l'aide et la position de l'erreur :

$F(2,24)$ = 3,19 ; p=.05 ; η^2_p=.21

La Figure 6-2 indique que les évaluations subjectives de charge cognitive faites avec *'Workload Profile'* ont eu une évolution dont la forme est proche de celle obtenue pour le score de rappel. La charge était plus faible lorsque les erreurs apparaissaient dès le premier dialogue, particulièrement pour la condition *'aides réparties'*. On remarque par ailleurs que dans la condition *'sans aide'* et lorsque les erreurs sont apparues à l'essai 2, les évaluations de charge ont eu un étalement important, ce qui indique des différences importantes au sein de ce groupe en ce qui concerne le « vécu subjectif » lié à ces erreurs.

B Indicateurs de réalisation de la tâche

La même procédure GLM a encore été répétée pour chacun des trois indicateurs de réalisation de la tâche présentés dans le Tableau 6-2. Chaque fois, les dialogues avec et sans erreurs ont été intégrés en tant que mesures répétées et les deux autres facteurs (*'condition d'aide'* et *'position du dialogue avec erreurs'*) ont été intégrés comme facteurs catégoriels. Pour chacun de ces trois indicateurs, les deux autres ont été inclus en co-variables.

Le Tableau 6-2 présente une lecture des résultats selon la *'condition d'aide'* et l'*essai avec ou sans erreur'*, mais qui ne tient pas compte de la *'position du dialogue avec erreur'*.

Tableau 6-2 : Durée, nombre de mots et de tours de parole pour les deux dialogues

Essai	Indicateur	**Sans aide**	**Aides groupées**	**Aides réparties**	Test de Levene
Dialogue sans erreur	***TDP***	4,52 (1,2)	4,82 (1,0)	5,82 (1,4)	*F(2,66)*=0,74; *p*=.47
	Mots	40,39 (11,9)	43,45 (19,5)	35,34 (15,8)	*F(2,66)*=2,11; *p*=.12
	Durée *(sec.)*	69,52 (9,9)	123,8 (21,9)	115,52 (14,7)	*F(2,66)*=1,90; *p*=.15
Dialogue avec erreurs	***TDP***	12,86 (4,7)	10,15 (2,5)	12,21 (4,0)	***F(2,66)*=9,25; *p*<.001**
	Mots	92,17 (44,1)	89,88 (39,8)	85,52 (38,6)	*F(2,66)*=0,14; *p*=.86
	Durée *(sec.)*	172 (37,3)	191,64 (43,7)	200,52 (49,9)	*F(2,66)*=1,51; *p*=.22

Les comparaisons ont révélé que l'essai *'avec ou sans erreur'* (facteur 2) n'a conduit à des différences significatives que pour la *'durée des dialogues'* (***F(1,59)* = 8,77 ; *p*<.01 ; η^2_p=.13**). Il y avait un **effet marginal** de ce facteur sur le *'nombre de tours de parole'* (*F(1,59)* = 3,22 ; *p*=.077 ; η^2_p=.05), mais aucun effet sur le *'nombre de mots'* (*F(1,59)* = 0,56 ; *p*=.45 ; η^2_p=.01). On lit pourtant des différences importantes dans le Tableau 6-2, mais ces différences ne tiennent pas compte des effets combinés des trois facteurs de l'expérience :

- Chacun des trois indicateurs dépendait significativement de la *'condition d'aide'* (facteur 1). Pour ce facteur, les indices restaient proches du seuil de significativité pour le *'nombre*

de tours de parole' (***F(2,59)* = 4,02 ; *p*<.05 ; η^2_p=.12**) et pour le *'nombre de mots'* (***F(2,59)* = 3,31 ; *p*<.05 ; η^2_p=.10**) ; et la différence était beaucoup plus marquée pour la *'durée des dialogues'* (***F(2,59)* = 24,47 ; *p*<.000 ; η^2_p=.47**) ;

- La *'position du dialogue avec erreur'* (facteur 3) a eu un effet significatif uniquement sur le *'nombre de tours de parole'* (***F(1,59)* = 5,73 ; *p*<.02 ; η^2_p=.08**). Dans toutes les conditions, les participants ont fait plus de *'tours de parole'* quand le *'dialogue avec erreurs'* apparaissait au deuxième essai. Ce facteur *'position de l'erreur'* n'a pas eu d'effet direct sur le *'nombre de mots'* (*F(1,59)* = 0,32 ; *p*=.57 ; η^2_p=.00) ni sur la *'durée des dialogues'* (*F(1,59)* = 0,67 ; *p*=.41 ; η^2_p=.01). Cependant, le *'nombre de mots'* dépendait d'une interaction entre *'position de l'erreur'* et *'essai avec ou sans erreur'* (***F(1,59)* = 4,40 ; *p*<.05 ; η^2_p=.07**). La différence entre les deux essais était moins importante quand les erreurs apparaissaient dans le second dialogue.

Dans le *'dialogue sans erreur'*, les participants ont fait plus de tours de parole pour la condition *'aides réparties'*, mais ils ont produits moins de mots. C'est dans la condition *'aides groupées'* qu'ils ont été le plus verbeux ; et cette condition a également donné lieu aux dialogues les plus longs. Quand les aides étaient *'groupées'*, les dialogues ont été plus longs de 54 secondes par rapport à la condition *'sans aide'*, soit précisément la durée des messages d'aide. Le prolongement n'était que de 45 secondes dans la condition *'aides réparties'*. Cela indique que dans cette condition les participants ont été plus efficaces.

Dans le *'dialogue avec erreurs de reconnaissance vocale'*, les différences sont moins marquées. On remarque surtout un nombre de tours de parole plus faible dans la condition *'aides groupées'* et une annulation des différences entre les trois conditions en ce qui concerne le *'nombre de mots'* produits. Les différences en ce qui concerne la *'durée des dialogues'* ont été ramenées, par rapport à la condition *'sans aide'*, à 20 secondes pour les *'aides groupées'* et à 28 secondes pour les *'aides réparties'*. Ces différences indiquent une meilleure performance des participants dans la correction des erreurs pour les conditions avec aides. Elles peuvent être expliquées en observant le comportement stratégique des participants au cours des dialogues (*Cf.* ci-dessous).

C Indicateurs comportementaux

Les résultats proposés concernant le comportement stratégique portent sur le moment de la prise de parole, notamment pour la correction des erreurs, sur le mode de correction des erreurs de reconnaissance vocale et sur le nombre de mots nécessaires à la correction en fonction du mode de correction.

- ***Moment de la prise de parole***

Les résultats du Tableau 6-3 présentent les moments des prises de parole dans la phase de choix. Cette étape est la phase centrale dans le dialogue avec ce service (*Cf.* Figure 5-3). Au cours du dialogue avec erreurs, les participants ont eu à prendre la parole trois fois dans

cette phase : pour corriger les deux premières erreurs, puis pour faire leur choix final. Le tableau présente la distribution des réponses relativement aux énoncés du système.

Tableau 6-3 : Moment de la prise de parole dans la phase de choix

Etape	*Prise de parole*	**Sans aide**	**Aides groupées**	**Aides réparties**	**X^2**
Correction erreur 1	*Dès l'écho*	2	12	13	***$X^2(4)$* = 27,68 ; *p* < .000**
	Pendant la liste	2	7	6	
	Après la relance	19	4	4	
Correction erreur 2	*Dès l'écho*	3	13	21	***$X^2(4)$* = 39,33 ; *p* < .000**
	Pendant la liste	2	6	1	
	Après la relance	18	4	1	
Choix final	*Pendant la liste*	10	17	21	***$X^2(2)$* = 12,73 ; *p* < .001**
	Après la relance	13	6	2	

On constate que les messages d'aide ont influencé les participants dans leurs comportements. Les participants qui ont reçu des aides ont eu un comportement beaucoup plus réactif que ceux qui n'en ont pas reçu. Notamment, dans la condition 'aides réparties' les participants ont eu tendance à prendre la parole très tôt. Au contraire, dans la condition 'sans aide' les participants sont restés passifs et ont eu tendance à attendre une question explicite du système (la relance) avant de prendre la parole.

- ***Mode de correction des erreurs de reconnaissance vocale***

Quand une erreur de reconnaissance vocale apparaissait, le participant pouvait dire « annuler » (étape 1) puis reformuler sa requête (étape 2). Il s'agissait alors d'une *'annulation'*. Sinon, il pouvait reformuler directement sa requête, en une seule étape. Il s'agissait alors d'une *'correction'*. Le Tableau 6-4 porte sur les dialogues *'avec erreurs'* et indique le nombre de participants ayant utilisé l'un ou l'autre de ces modes pour les deux erreurs successives.

Tableau 6-4 : Mode de correction des erreurs de reconnaissance vocale

	Mode de correction	**Sans aide**	**Aides groupées**	**Aides réparties**	**X^2**
Correction erreur 1	***Annulation***	15	13	12	$X^2(2)$ = 0,83 ; *N.S.*
	Correction	8	10	11	
Correction erreur 2	***Annulation***	10	17	21	***$X^2(2)$* = 12,73 ; *p* < .01**
	Correction	13	6	2	

La première erreur a donné lieu à des nombres d'*annulations* et de *corrections* sensiblement équivalents entre les différentes conditions. La distribution la plus déséquilibrée a été trouvée dans le groupe *'sans aide'* dans lequel les *'annulations'* étaient légèrement plus nombreuses. Dans les deux autres groupes, les effectifs étaient équivalents. Après la seconde erreur de reconnaissance vocale, les participants qui ont reçu des aides ont eu tendance à privilégier

l'*annulation'* alors que ceux qui n'ont pas entendu les aides ont privilégié la *'correction'*. Ainsi, les aides ont influencé les choix stratégiques des participants.

- ***Nombre de mots nécessaires à la correction***

Les effectifs relativement équilibrés entre les deux modes de correction et pour la première erreur, autorisent à utiliser cette variable comme facteur pour une ANOVA. La Figure 6-3 présente le nombre de mots utilisés pour corriger la première erreur en fonction du *mode de correction* choisi et du *moment de la prise de parole*.

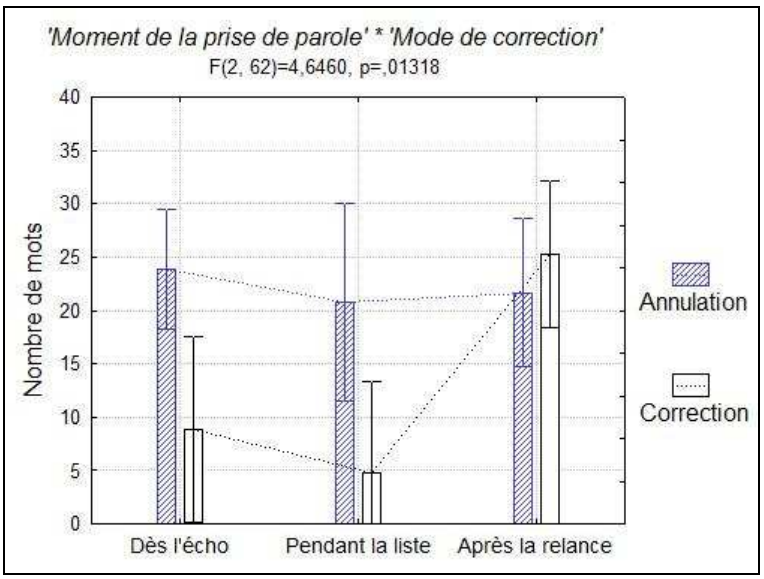


Figure 6-3 : Nombre de mots pour corriger la 1ère erreur

Le test de Levene révèle des distributions homogènes pour le nombre de mots :
F(1,67)=0,01 ; *p*=.89

Le mode de correction de l'erreur a eu un effet sur le nombre de mots :
***F(1,62)*=7,58 ; *p*<.01 ; η^2_p=.10**

Le moment de la correction a également eu un effet sur le nombre de mots :
***F(2,62)*=4,16 ; *p*<.05 ; η^2_p=.11**

Et il y a eu un effet d'interaction entre ces deux variables :
***F(2,62)*=4,64 ; *p*<.02 ; η^2_p=.13**

On constate que les participants qui ont fait une correction directe *'dès l'écho'* ou *'pendant la liste'* ont fait leur correction en moins de 10 mots en moyenne, alors que dans toutes les autres conditions, plus de 20 mots (en moyenne) ont été nécessaires. Ainsi, il semble exister une dépendance forte entre cet indicateur (le nombre de mots) et le comportement stratégique mis en place (le mode de correction).

6.1.4 Conclusion

Ces résultats ont permis de constater que le surplus d'information lié aux messages d'aide ne provoquait pas de gêne concernant la mémorisation des informations cible (les *'rendez-vous'*). Le rappel a été sensible surtout à la position du dialogue avec erreurs et au fait de distribuer les aides dans le dialogue. Par ailleurs, les évaluations subjectives de charge cognitive n'ont pas révélé d'augmentation quand des informations supplémentaires étaient présentes (dues *aux aides* et/ou *aux erreurs*). Au contraire, quand les erreurs étaient présentes dès le premier essai les participants ont évalué une charge cognitive plus faible.

Du point de vue théorique, ces résultats permettent de considérer que les aides n'ont pas surchargé la mémoire de travail des participants puisque leur présence n'a pas, en elle-même,

entrainé d'effet négatif sur le rappel des informations cible. Ce résultat contredit la première des trois hypothèses formulées à partir de la théorie de la charge cognitive : *la charge liée au traitement des aides n'a pas interféré avec la charge liée à l'acquisition des informations sur le rendez-vous recherché*. En revanche, il est possible de considérer que les hypothèses 2 et 3, formulées à partir de cette théorie, sont vérifiées. En effet, pour les participants qui n'ont rencontré des erreurs qu'au second essai, les aides réparties ont été bénéfiques. Elles ont donné lieu à un meilleur score de rappel des informations cible (hypothèse 2). Mais ce bénéfice a été perdu pour les participants qui ont rencontré des erreurs dès le premier essai (hypothèse 3). Cependant, pour considérer que ces hypothèses sont vérifiées, il est nécessaire de supposer que la surcharge cognitive liée au traitement des erreurs a été dépensée quand les erreurs apparaissaient dès le premier essai, mais pas quand elles apparaissent au second essai. Dans ce cas, les connaissances procédurales acquises par le participant dans le premier dialogue lui auraient permis de traiter les erreurs dans le second dialogue sans surcoût cognitif. Ce point n'était pas directement formulé à partir de la théorie de la charge cognitive, mais il est cohérent avec les prédictions qu'elle permet. Ainsi, selon ce modèle, après un essai d'entraînement, la charge cognitive liée au traitement des erreurs était suffisamment faible pour ne pas interférer avec la charge cognitive liée à l'acquisition des informations. Cette explication est basée sur la notion de concurrence pour les ressources cognitives.

Les trois hypothèses proposées à partir du *point de vue pragmatique* sont vérifiées. La première hypothèse prédisait que le traitement des aides n'entrait pas en concurrence avec le traitement des informations sur les rendez-vous, mais qu'au contraire, leur prise en compte était susceptible de faciliter le processus interactif. En effet, la présence des aides n'a pas eu d'impact sur le rappel ni sur les évaluations subjectives de charge cognitive. De plus, un effet bénéfique sur le rappel a été obtenu pour les aides réparties. Conformément à ce résultat, l'hypothèse 2 prédisait que les aides réparties correspondent au fonctionnement du modèle de contribution, dans lequel les différentes contributions font l'objet de validations successives. De même, l'hypothèse 3 indiquait que les participants ayant entendu les aides s'adapteraient plus efficacement. Les résultats sur les indicateurs comportementaux et sur les indicateurs de réalisation de la tâche vont dans ce sens. Les participants qui ont entendu les aides ont été plus réactifs (*Cf.* Tableau 6-3). Ils se sont orientés vers une stratégie particulière (l'*annulation, Cf.* Tableau 6-4). Dans la condition *'aides réparties'* les participants ont fait plus de tours de parole et ils ont utilisé moins de mots (*Cf.* Tableau 6-2). Ils ont ainsi exercé un contrôle du dialogue plus efficace, sur un mode conversationnel.

Ce mode de contrôle conversationnel a permis à ces participants de focaliser leur attention sur la finalité du dialogue (la prise de rendez-vous). C'est probablement pour cette raison qu'un meilleur rappel a été obtenu dans la condition *'aide répartie'*. Mais la disparition de cet effet quand les erreurs de reconnaissance vocale apparaissaient dès le premier dialogue suggère qu'il y a bien eu une *concurrence* entre les traitements liés à la *correction des erreurs* et les traitements liés à l'*acquisition des informations*. Cette concurrence est fondamentale dans la théorie de la charge cognitive. La notion de *capacité limitée* pourrait

être utilisée pour expliquer ce résultat. Cependant, la concurrence pour les ressources suppose que les évaluations de charge cognitive auraient dû être plus élevées dans cette condition. Or, c'est le contraire qui a été observé. Quand les erreurs sont apparues dès le premier dialogue la charge était plus faible. Dans le cadre de la théorie pragmatique, il est possible d'avancer que ce sont les inférences des participants qui ont eu un effet (sur le rôle des inférences dans le dialogue, voir Grice, 1957; 1969; 1975). En effet, les erreurs rencontrées dès la première expérience avec le système semblent avoir démotivé les participants. Ils se sont moins investis (charge faible) et ont été moins attentifs (rappel faible). Ainsi, dans une certaine mesure, il semble préférable de maintenir une charge cognitive suffisamment importante pour soutenir la mémorisation. – Dans la condition la plus favorable, l'évaluation moyenne était d'environ 50 (sur une échelle de 100), ce qui indique une charge médiante pour cette condition. Rien n'indique qu'il faille *surcharger* l'utilisateur. – Il serait alors préférable de chercher à *motiver l'utilisateur* pour le stimuler à activer ses ressources cognitives plutôt que de chercher à *ne pas le surcharger*.

Le temps perdu par la diffusion des messages d'aide (54 secondes) n'a pas été compensé dans le *'dialogue sans erreur'* et il n'a été que partiellement compensé dans le *'dialogue avec erreurs'* (*Cf.* Tableau 6-2). L'implémentation du système, liée au protocole en *Magicien d'Oz*, peut expliquer en partie ce résultat. La durée moyenne de 200 secondes dans le *'dialogue avec erreurs' et* pour la condition *'aides réparties'* est liée à la simplicité de fonctionnement du système. Dans cette condition, les participants ont entendu trois fois l'aide à la correction d'erreur (16 secondes) et l'aide à la navigation dans la liste (15 secondes) car ces messages étaient diffusés à chaque nouvelle requête (après l'*écho* de la requête et avant la présentation des *réponses* correspondantes, *Cf.* Figure 5-3). La suppression de la troisième diffusion de ces messages (facile à implémenter dans un système en fonctionnement réel) aurait conduit à des dialogues plus courts que dans la condition *'sans aide'*. Indépendamment de cette dernière remarque, et plus encore si on l'accepte, la condition *'aides réparties'* semble la plus favorable à l'utilisation correcte et performante du système. La compensation partielle obtenue dans les conditions avec aides est liée à la réactivité plus importante des participants. Face aux erreurs de reconnaissance vocale, ils ont eu une tendance beaucoup plus forte à réagir dans l'instant (*'dès l'écho'*, *Cf.* Tableau 6-3). Cette réactivité n'a été notée, dans la condition *'sans aide'*, dans aucune des trois *étapes* reportées dans le Tableau 6-3. Ainsi, les informations exposées aux participants dans ces messages leur ont permis de comprendre comment se comporter face au système ; comportement qu'ils ont immédiatement mis en place.

Enfin, les résultats concernant le *nombre de mots* nécessaires à la correction (en fonction du *mode de correction* choisi) présente un intérêt qui va dans le sens de l'optimisation du fonctionnement des systèmes de dialogue. En effet, il est possible d'accroître les capacités de compréhension des systèmes de dialogue en utilisant des critères sociocognitifs pour qualifier les interventions de l'utilisateur. Parmi les critères qui peuvent être proposés dans le modèle sociocognitif (Clark, 1996) certains sont indépendants des analyses verbales et grammaticales faites à partir de la reconnaissance vocale (*e.g.* durée des énoncés, débit de

parole, intensité de la voix, etc.). On sait par ailleurs que les erreurs de reconnaissance vocale sont fréquentes. Or, l'interprétation du système est basée sur la transcription mot à mot des énoncés de l'utilisateur et sur des regroupements syntaxiques (groupes nominaux, groupes verbaux, etc.) qu'elle permet. Les erreurs de reconnaissance vocale (substitutions, rejets, etc.) apparaissent au cours de la transcription. D'autres types d'erreurs (d'ordre sémantique et fonctionnel) apparaissent au cours de l'interprétation. Le résultat présenté ici montre qu'il est possible (après apprentissage) de faire des hypothèses sur ce qui est dit, indépendamment de cette transcription. Si le système dispose d'un modèle de la tâche suffisant, il peut apprendre à identifier les formes comportementales les plus fréquentes dans les différentes phases de cette tâche. La forme de surface des commandes de l'utilisateur (par exemple, leur durée ou le nombre de mots qu'elles contiennent) peut en effet permettre de produire d'autres hypothèses que celles de l'analyse grammaticale de façon à les confronter entre elles pour augmenter la probabilité d'obtenir une interprétation correcte. De cette manière, il est possible d'améliorer la performance d'interprétation des commandes sans augmenter, au sens strict, la performance de reconnaissance vocale. Du point de vue technique, ce résultat incite à développer le parallélisme des traitements dans les systèmes. Du point de vue psycho-ergonomique, il incite à favoriser le développement de formalismes capables de rendre compte du déploiement stratégique de l'activité. Dans un cas comme dans l'autre, c'est une *'théorie de l'action'* qui permet d'en rendre compte.

6.2 Expérience 2 : Effet de la verbosité du système

6.2.1 Objectifs et hypothèses

Les énoncés du système sont consommateurs de temps. Les études montrent (Bubb-Lewis & Scerbo, 2002; Walker et al., 2004), tout comme l'expérience 1 (ci-avant), que les informations fournies aux utilisateurs sont souvent utiles dans l'interaction, mais qu'elles peuvent engendrer des coûts temporels qui dégradent la performance. De ce fait, il peut sembler judicieux de chercher à raccourcir les énoncés du système pour gagner du temps.

Par ailleurs, le processus de conception des énoncés au cours des projets implique que des concepteurs travaillent sur le contenu et la forme des énoncés. Le but est d'exposer clairement à l'utilisateur toutes les informations dont il a besoin au cours du processus dialogique. Le travail de ces énoncés sous une forme écrite incite les concepteurs à produire une syntaxe correcte qui peut parfois avoir pour conséquence de produire des messages longs. Or, on peut se demander si ces formulations syntaxiquement correctes sont souhaitables, ou du moins, si elles sont nécessaires ; et quelles sont les conséquences de leur suppression. Peut-on guider correctement l'utilisateur en limitant la longueur des énoncés système ? Quel est l'impact sur l'exploration des réponses ? Quel est l'impact en mémoire ? Quel est l'impact sur le comportement verbal de l'utilisateur ?

L'objectif est de vérifier si un style télégraphique (qui serait plus facile à faire produire automatiquement par un robot) est suffisant pour une gestion correcte des interactions et si la performance (du point de vue cognitif et du point de vue de l'efficacité du dialogue) en est modifiée. On s'intéresse notamment aux stratégies d'exploration et aux prises et cession de parole. Le service qui doit permettre de le tester contient plus de données, ce qui permet d'observer la manière dont les participants s'y prennent pour interrompre le système et prendre l'initiative.

Dans la perspective pragmatique, la modification des énoncés pourrait d'abord modifier les règles de prise de parole et provoquer des effets sur les parcours des participants dans les dialogues, modifiant ainsi la structure des interactions.

Par ailleurs, le fait d'accélérer la *vitesse de défilement* des informations est susceptible d'accroître la difficulté perceptive. De plus, la *structure syntaxique* fournit une aide dans le processus d'organisation des informations. Sa suppression pourrait perturber ce processus et dégrader la compréhension des énoncés du système. La combinaison de ces deux facteurs est susceptible de faire décroître la performance de rappel.

6.2.2 Méthode

A Participants

Le nombre final de participants était de 45 (9 hommes, 36 femmes). La moyenne d'âge du groupe était de 22,34 ans (*é.t.* = 4,92 ; *min.* = 19,25 ; *max.* = 40,25). Les participants étaient des étudiants en psychologie à l'Université Rennes 2 en Licence – deuxième année. Ils participaient à l'expérience dans le cadre des enseignements en psychologie expérimentale, en tant qu'activité obligatoire associée au cours.

Le niveau d'expérience en informatique des participants était moyen. Ils ont déclaré passer en moyenne 12,15 heures par mois à utiliser un ordinateur (*é.t.* = 12,8 ; *min.* = 1 ; *max.* = 70) dont 6,46 sur internet (*é.t.* = 7,20 ; *min.* = 1 ; *max.* = 30). Pour comparaison, ils ont déclaré passer 11,68 heures par mois à regarder la télévision (*é.t.* = 8,63 ; *min.* = 0 ; *max.* = 35) et . 7,68 heures par mois à lire (revues et/ou littérature : *é.t.* = 6,46 ; *min.* = 0 ; *max.* = 29).

B Matériel

Le système utilisé pour cette expérience était une simulation du service *'Cinéliste'* (*Cf.* Figure 5-4) développée et mise en place selon le dispositif expérimental présenté au chapitre 5. Ce service correspond à un dialogue en quatre phases (*'Requête'–'Choix'–'Détail'–'Confirmation'*) dédiées au choix d'un film. La phase de *'détail'* supplémentaire donne accès à une quantité d'informations plus importante concernant la référence (le film choisi) que dans le service *'Santiago'* utilisé dans l'expérience 1. L'objectif était d'évaluer le rappel de ces informations. Dans la mesure où la quantité d'information était plus importante, il était supposé que l'évaluation serait plus discriminante.

C Procédure

Le participant était accueilli individuellement. Il était conduit dans une pièce calme où était installé le matériel expérimental. L'expérimentateur lui présentait brièvement, à l'oral, le service *'Cinéliste'* et le matériel qu'il s'apprêtait à utiliser pour réaliser les appels. Il lui remettait ensuite deux documents (*Cf.* Annexes, p. 260) :

(1) Le premier comprenait *trois feuilles de papier* dédiées au rappel des informations concernant les films recherchés. Ces feuilles proposaient des tableaux dont la colonne de droite était vide. La colonne de gauche indiquait les noms des champs à remplir (*'Titre'*, *'Genre'*, *'Nationalité'*, *'Année'*, *'Acteur 1'*, *'Acteur 2'*, *'Synopsis'*). La première feuille, pour le scénario de familiarisation, comprenait un seul tableau. La seconde feuille, pour l'essai simple, comprenait un seul tableau, pour le film à identifier. La troisième feuille, pour l'essai complexe, comprenait deux tableaux, pour les deux films qui devaient être identifiés. Du fait de la présentation de ces feuilles de rappel au participant, la consigne de rappel était explicite dans cette expérience ;

(2) L'autre document présentait les *consignes liées à l'évaluation de la charge cognitive*. Il incluait une page de présentation générale du questionnaire *'Workload Profile'* et une page qui présentait chacune des dimensions à évaluer, selon les prescriptions des concepteurs de ce questionnaire (Tsang & Velasquez, 1996).

Après avoir pris connaissance de ces documents, le participant commençait la procédure expérimentale sur l'ordinateur dédié. La procédure était organisée par le programme de test qui présentait le questionnaire d'évaluation de la charge subjective suite à chaque dialogue et un questionnaire final pour demander au participant d'évaluer l'utilisabilité du système et pour obtenir les renseignements généraux le concernant.

D Protocole expérimental

Deux variables indépendantes ont été manipulées dans cette expérience.

- ***Facteur 1*: Style d'énoncé**

Trois conditions ont été évaluées, correspondant à trois *« niveaux de verbosité »* du système :

1. ***'Style normal'*** : Les énoncés du système étaient conçus selon une syntaxe normale, correspondant au style adopté habituellement dans la conception de services vocaux pour le grand public. Les phrases étaient syntaxiquement complètes et correctement structurées ;
2. ***'Style condensé'*** : Dans cette condition, le nombre de mots prononcés par le système a été divisé par deux. La structure syntaxique des phrases a été supprimée. Par exemple, l'énoncé : *« Bienvenue sur le service Cinéliste »* a été remplacé par : *« Service Cinéliste »*. Certaines mises en contexte ont également été éliminées. Par exemple, l'énoncé : *« Vous pouvez demander la lecture du film, dites lecture ; réécouter les détails, dites réécouter ; ou bien retourner à la liste, dites retour. Pour passer à un autre film, dites suivant ou précédent. »* a été remplacé par : *« Commandes acceptées : LECTURE, REECOUTER, SUIVANT, PRECEDENT, RETOUR. »* ;
3. ***'Style espacé'*** : Dans une troisième version, les énoncés de *'style condensé'* (*condition 2*) ont été réutilisés. Dans la mesure où ces énoncés sont plus courts, des temps de silence y ont été ajoutés de façon à ce que leur diffusion prenne autant de temps que celui nécessaire à la diffusion des énoncés de style normal (*condition 1*).

Il s'agit d'une *variable intergroupe*. Trois groupes ont été constitués, correspondant aux trois conditions. Chaque participant a rencontré une seule de ces conditions au cours de l'expérience.

- ***Facteur 2* : Essai**

Chaque participant a réalisé trois dialogues. Le premier dialogue correspondait à un essai de familiarisation avec le système et n'est pas intégré aux résultats expérimentaux. Le second dialogue avait une consigne simple, le troisième avait une consigne complexe :

- ***'Essai simple'*** : Le participant avait pour consigne d'identifier un film et d'en demander la lecture. La question était : « Dans quel film Madame Baines meurt-elle ? » ;
- ***'Essai complexe'*** : Le participant avait pour consigne d'identifier deux films et de demander la lecture de l'un des deux. La question était : *« Identifiez le film le plus ancien et le plus récent parmi les films de Fritz Lang et demandez la lecture de l'un d'eux. »*

Il s'agit d'une *variable intragroupe*. Chaque participant rencontrait les deux conditions au cours de deux essais successifs.

E Mesures dépendantes

Les mesures utilisées dans la présentation des résultats sont indiquées dans le Tableau 6-5. Lorsque des explications sur la signification des différents indicateurs sont nécessaires, elles sont précisées lors de la présentation des résultats.

Tableau 6-5 : Indicateurs utilisés pour la présentation des résultats de l'expérience 2

Rappel	Nombre d'items (mots) correctement rappelés parmi 30 (informations sur le film : *'titre'*, *'année', 'genre', 'nationalité', 'acteur 1', 'acteur 2' ;* et histoire du film : *'synopsis'*).
Charge cognitive subjective	Evaluée avec le questionnaire *'Workload Profile'*.
Nombre de tours de parole (TDP)	Nombre total de prises de parole par le participant au cours du dialogue.
Nombre de mots	Nombre total de mots prononcés par le participant au cours du dialogue.
Durée	Durée totale du dialogue (en secondes).
Nombre d'écoutes des informations	Nombre de consultations du ou des film(s) identifié(s) par le participant au cours du dialogue.
Mode d'exploration de la liste	(1) Direct, (2) Indirect. (Voir détail dans la présentation des résultats).
Moment de la prise de parole	Catégorisé relativement à l'énoncé du système : (1) *'Pendant les réponses'*, (2) *'Pendant le menu'*, (3) *'Après la relance'*.
Taux de recouvrement	Entre parole du participant et parole du système. Ce taux correspond au pourcentage de prises de parole du participant au cours desquelles un temps de recouvrement est apparu.

6.2.3 Résultats

Le corpus final est composé de 90 dialogues au cours desquels 1675 commandes des participants ont été prises en compte pour un total de 2056 mots. Du fait de la structure de menu, les participants ont utilisé de préférence des commandes en un seul mot permettant de « naviguer dans les données ». Le nombre moyen de tours de parole (soit le nombre

d'énoncés prononcés pour commander le système) était de 14,07 pour le scénario simple et de 23,16 pour le scénario complexe.

Ces dialogues ont été transcrits mot à mot, incluant les hésitations et autres *'disfluences'* de la parole des participants. En complément à ces transcriptions mot à mot : (1) le « moment de la prise de parole » du participant était indiqué relativement à la structure des énoncés du système ; et (2) l'apparition de recouvrements entre la parole de l'utilisateur et celle du système était également indiquée. Les résultats présentés ne différencient pas les recouvrements en début de parole de l'utilisateur (l'utilisateur coupe le système) ou en fin de parole de l'utilisateur (le système coupe l'utilisateur). Les analyses statistiques réalisées sur les données sont précisées relativement à chaque indicateur présenté.

A Indicateurs cognitifs généraux

La performance de rappel est évaluée par le nombre d'items rappelés pour chaque film. Parmi ces items, 9 concernent les informations sur le film (*'titre'* : deux items ; *'année'* : un item ; *'genre'* : un item ; *'nationalité'* : un item ; *'acteur 1'* : deux items ; *'acteur 2'* : deux items) et 21 portent sur le synopsis (résumé de l'histoire en 21 ou 22 mots selon les films). Le total est de 30 items par film.

Les résultats ont été analysés avec une procédure GLM utilisant le *'style d'énoncé'* (facteur 1) comme facteur catégoriel et la variable *'essai'* (facteur 2) en tant que mesure répétée. Le *'nombre de tours de parole'*, le *'nombre de mots'* et la *'durée des dialogues'* expérimentaux ont été inclus en co-variables. Cette procédure a été exécutée séparément pour le rappel des informations et pour les évaluations subjectives de charge cognitive.

Tableau 6-6 : Rappel et charge cognitive subjective pour les deux dialogues

Essai	Indicateur	**Normal**	**Condensé**	**Espacé**	Test de Levene
Simple	***Rappel***	17,2 (5,7)	17,6 (6,9)	14,2 (6,6)	*F(2,42)*=0,19; *p*=.82
	Charge (WP)	33,2 (12,8)	35,0 (10,8)	34,7 (12,6)	*F(2,42)*=0,45; *p*=.63
Complexe	***Rappel***	11,4 (3,7)	10,7 (5,9)	11,2 (4,7)	*F(2,42)*=1,23; *p*=.30
	Charge (WP)	35,1 (17,2)	36,5 (10,7)	36,8 (12,3)	*F(2,42)*=1,70; *p*=.19

Le *'style d'énoncé'* n'a pas eu d'effet sur la performance de rappel (*F(2,36)*=0,63 ; *p*=.53 ; η^2_p=.03), ni sur les évaluations de charge cognitive (*F(2,36)*=0,00 ; *p*=.99 ; η^2_p=.00).

Le facteur *'essai'* a eu un effet sur la charge cognitive (***F(1,36)*=5,48 ; *p*<.05 ; η^2_p=.13**). La charge était plus faible à l'essai simple. En revanche, aucun effet sur le rappel (*F(1,36)*=0,88 ; *p*=.35 ; η^2_p=.02) n'est apparu.

Il n'y avait pas d'effet d'interaction entre les facteurs pour le rappel (*F(2,36)*=1,54 ; *p*=.22 ; η^2_p=.07) ni pour les évaluations de charge cognitives (*F(2,36)*=0,31 ; *p*=.73 ; η^2_p=.01).

Ainsi, le style d'énoncé n'a pas eu d'effet sur les indicateurs cognitifs généraux. Il faut cependant rappeler que le comportement des participants était libre et que le test de rappel faisait parti des consignes de l'expérience. Pour répondre à la consigne, les participants ont réécouté plusieurs fois les informations sur les films. Le nombre d'écoutes est présenté plus loin, et une interprétation en est proposée, parmi les indicateurs de comportement stratégique.

B Indicateurs de réalisation de la tâche

La même procédure GLM (facteur 1 en prédicteur catégoriel, facteur 2 en mesure répétée) a été utilisée pour le *'nombre de tours de parole'*, le *'nombre de mots'* prononcés par le participant et la *'durée du dialogue'*. Pour chacun de ces trois indicateurs, les deux autres ont été inclus en co-variables.

Le *'style d'énoncé'* n'a eu d'effet sur aucun de ces trois indicateurs : *nombre de tours de parole* (*F(2,38)*=1,32 ; *p*=.27 ; η^2_p=.06), *nombre de mots* (*F(2,38)*=1,66 ; *p*=.20 ; η^2_p=.08) et *durée des dialogues* (*F(2,38)*=2,15 ; *p*=.13 ; η^2_p=.00).

L'*essai'* a eu un effet sur le nombre de tours de parole (***F(1,38)*=4,56 ; *p*<.05 ; η^2_p=.10**), plus important à l'essai complexe ; mais il n'a eu aucun effet sur le nombre de mots prononcés (*F(1,38)*=0,35 ; *p*=.55 ; η^2_p=.00) ni sur la durée des dialogues (*F(1,38)*=2,77 ; *p*=.10 ; η^2_p=.06).

Un **effet d'interaction marginal** est apparu entre les deux facteurs expérimentaux pour le nombre de mots prononcés (*F(2,38)*=3,02 ; *p*=.06 ; η^2_p=.13). A l'essai simple, le nombre de mots était plus faible dans la condition de *'style espacé'* ; mais cette différence n'apparaissait plus à l'essai complexe. Cet effet n'est pas apparu pour le nombre de tours de parole (*F(2,38)*=1,48 ; *p*=.23 ; η^2_p=.07), ni pour la durée (*F(2,38)*=1,43 ; *p*=.25 ; η^2_p=.07).

Tableau 6-7 : Durée, nombre de mots et de tours de parole pour les deux dialogues

Essai	Indicateur	**Normal**	**Condensé**	**Espacé**	Test de Levene
Simple	***TDP***	15,6 (4,40)	13,8 (4,70)	12,8 (3,54)	*F(2,42)*=0,38; *p*=.68
	Mots	20,8 (7,72)	21,9 (15,6)	14,6 (4,77)	*F(2,42)*=2,89; *p*=.06
	Durée *(sec.)*	419,8 (118)	322,4 (106)	360,6 (89)	*F(2,42)*=0,44; *p*=.64
Complexe	***TDP***	24,2 (6,78)	22,4 (7,19)	22,8 (6,29)	*F(2,42)*=0,07; *p*=.92
	Mots	27,4 (7,16)	27,0 (11,7)	25,2 (8,88)	*F(2,42)*=1,10; *p*=.33
	Durée *(sec.)*	474,0 (113)	453,9 (177)	533,9 (167)	*F(2,42)*=1,52; *p*=.22

Ainsi, le style d'énoncé a eu peu d'effet sur les indicateurs de réalisation de la tâche. La légère baisse du nombre mots dans le dialogue simple pourrait révéler un comportement plus efficace, ou plus orienté vers la tâche, dans la condition de *'style espacé'*.

C Indicateurs comportementaux

Les indicateurs de comportement stratégique des utilisateurs présentés ici sont : le *'nombre d'écoutes'* des informations sur les films, le *'mode d'exploration'* de la liste des réponses, le *'moment de la prise de parole'* par les participants et le *'taux de parole avec recouvrement'*. Ces indicateurs renvoient aux décisions conscientes et aux automatismes mis en place par les participants qui ont une influence sur le déroulement de l'interaction.

- ***Nombre d'écoutes des informations sur les films***

Le *'nombre d'écoute des informations sur les films*' inclut la première écoute – lors de la recherche du (ou des) film(s) – et les écoutes suivantes, dédiées à la mémorisation du contenu. Pour l'essai complexe, le nombre d'écoutes présenté est une moyenne du nombre total d'écoute des deux films à identifier.

Ce nombre relève d'une décision consciente, de la part des participants, liée à l'apprentissage du contenu présenté. On peut l'interpréter (1) comme un indicateur de l'effort nécessaire à l'apprentissage : si ce nombre augmente c'est que le contenu est difficile à apprendre. On peut également l'interpréter (2) comme un indicateur de la facilité de navigation : si ce nombre augmente c'est que le participant trouve facile de réécouter les informations. Ces deux interprétations ne sont pas nécessairement contradictoires.

La même procédure GLM que celle utilisée pour le rappel et la charge cognitive a été utilisée pour le *'nombre d'écoutes'* des informations sur les films.

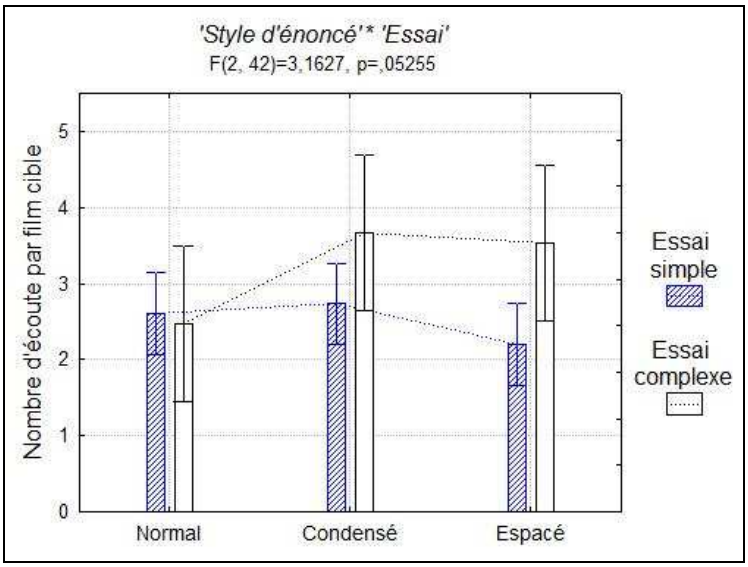


Figure 6-4 : Nombre d'écoutes des films cible

Les tests de Levene révèlent des distributions homogènes :
simple : F(2,42)=0,09; *p*=.90
complexe : *F(2,42)*=1,78; *p*=.18

Le *'style d'énoncé'* n'a pas eu d'effet sur le nombre d'écoutes :
F(2,42)=0,93 ; *p*=.40 ; η^2_p=.04

Le facteur *'essai'* a eu un effet significatif :
***F(1,42)*=8,34 ; *p*<.01 ; η^2_p=.16**

Un **effet d'interaction marginal** est apparu entre les deux facteurs :
(*F(2,42)*=3,16 ; *p*=.05 ; η^2_p=.13

On lit sur le graphique de la Figure 6-4 que le facteur *'essai'* a eu un effet sur le nombre d'écoutes des films dans les conditions de *'style condensé'* et de *'style espacé'*. Il n'y avait aucune différence entre les deux essais dans la condition de *'style normal'*. Ainsi, ce résultat semble indiquer que des constructions syntaxiques correctes et complètes permettraient une mémorisation plus rapide des contenus. Mais à l'inverse, les deux autres styles d'énoncés se sont différenciés par une plus grande facilité à réécouter les informations.

- ***Mode d'exploration***

Le mode d'exploration décrit le cheminement du participant parmi les phases du dialogue lors de l'exploration des données (*Cf.* Figure 5-4) :

- ***« Mode direct »*** : Le participant, (1) fait sa requête (phase de requête), (2) demande la consultation du premier film (phase de choix), (3) passe au(x) film(s) suivant(s) par ordre croissant jusqu'à identification du film voulu (phase de détail) et (4) réécoute les informations sur le film puis demande la lecture (phase de détail, puis de confirmation). Pour la dialogue complexe : (3') tous les films sont consultés par le participant pour identifier les films recherchés et (4') le participant réécoute les informations sur les deux films qu'il a identifiés.
- ***« Mode indirect »*** : Cette catégorie regroupe tous les autres modes d'exploration. Certains participants retournent à la liste des réponses (phase de choix) suite à la consultation du détail de chaque film (stratégie par ***« balayage »*** : *'film 1 – retour – film 2 – retour – film n – retour'*). D'autres participants tentent des films pris au hasard, sans tenir compte de leur rangement parmi les réponses du système (stratégie par ***« picorage »***). Par ailleurs, dans le dialogue complexe, certains participants ont eu tendance à réécouter les informations sur certains films après avoir identifié les films voulus pour contrôler l'absence d'erreur (stratégie avec ***« contrôle »***). Certains ont utilisé successivement ces différents modes d'exploration, notamment dans le dialogue complexe.

Tableau 6-8 : Modes d'exploration de la liste de films

Essai	Mode d'exploration	**Normal**	**Condensé**	**Espacé**	**X^2**
Simple	***direct***	6	8	12	$X^2(2)$ = 5,10 ; *N.S.*
	indirect	9	7	3	
Complexe	***direct***	10	8	13	$X^2(2)$ = 3,94 ; *N.S.*
	indirect	5	7	2	

On voit dans le Tableau 6-8 que le style d'énoncé a eu peu d'influence sur le mode d'exploration utilisé par les participants. Cependant, les effectifs sont déséquilibrés dans la condition de style *'espacé'*.

Dans les analyses statistiques présentées ci-dessous, le mode d'exploration utilisé par les participants est pris en compte en tant que prédicteur catégoriel.

- ***Moment de la prise de parole***

La description du moment de la prise de parole dans les dialogues expérimentaux donne lieu à des données complexes. Lors de l'exploration des données, les participants devaient faire des allers-retours entre les différentes phases du dialogue, de sorte que chacun a pris la parole un nombre variable de fois dans chaque phase. La *'phase de détail'* était la plus consultée (moy. = 8,26 ; é.t. = 2,27 ; min. = 2 ; max. = 12, pour l'essai simple). Pendant cette phase, le participant pouvait prendre la parole : (1) *'pendant les réponses'*, soit pendant que

le système présentait les informations sur le film en cours ; (2) *'pendant le menu'*, soit après la consultation de toutes les informations sur le film, mais avant que le système ne lui ait donné la parole ; ou (3) *'après la relance'* du système, soit après que le système lui ait donné la parole. Le Tableau 6-9 porte sur la *'phase de détail'* dans l'essai simple. Il présente la distribution moyenne, dans cette phase, des pourcentages de prises de parole aux différents moments. Chaque colonne a un total de 100.

On lit dans ce tableau que le *'style condensé'* provoque des réponses plus tardives des participants. Avec ce style d'énoncés 60% des réponses étaient faites après la relance, contre moins de 30% dans la condition de *'style normal'*. Les participants étaient donc plus patients avec le *'style condensé'*.

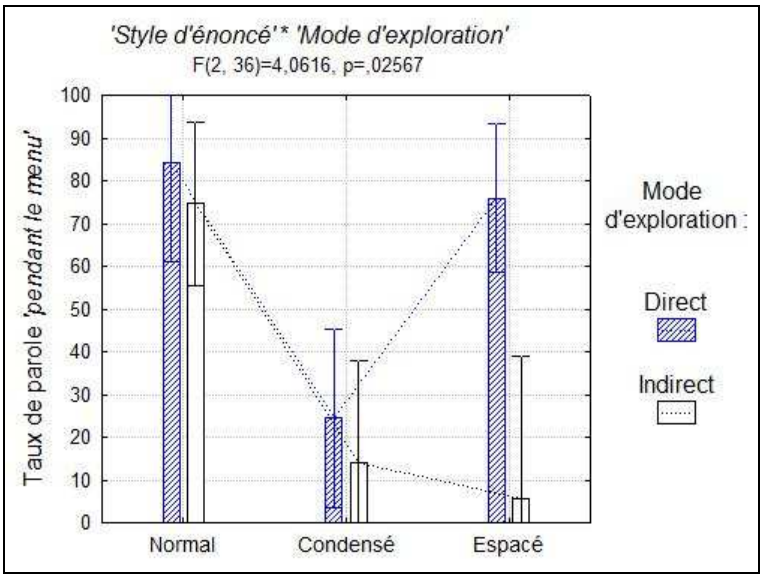
Tableau 6-9 : Distributions des moments de prises de parole dans la *'phase de détail'*

Moment	**Normal**	**Condensé**	**Espacé**	Test de Levene
Pendant les *'réponses'*	3,4 (6,6)	0,7 (2,8)	6,0 (12,2)	***F(5,39)*=5,40; *p*<.00**
Pendant le menu (*'relance'*)	68,4 (38,4)	39,2 (37,5)	52,2 (34,5)	*F(5,39)*=0,82; *p*=.53
Après la *'relance'*	28,0 (39,7)	60,0 (38,4)	41,7 (37,7)	*F(5,39)*=0,67; *p*=.64

Afin de faciliter la lecture des analyses statistiques, seuls les pourcentages correspondants aux réponses *'pendant le menu'* sont analysés ci-dessous.

- ***Prise de parole 'Pendant le menu'***

Les pourcentages de réponses *'pendant le menu'* ont été analysés avec une procédure GLM prenant le *'style d'énoncé'* et le *'mode d'exploration'* comme prédicteurs catégoriels. Le *'nombre de tours de parole'*, le *'nombre de mots'* et la *'durée des dialogues'* ont été utilisés en tant que co-variables. La Figure 6-5 présente le graphique obtenu.



Le style d'énoncé a eu un effet sur le taux de prises de parole pendant le menu :

***F(2,36)*=13,7 ; *p*<.00 ; η^2_p=.43**

Le mode d'exploration utilisé a également eu un effet :

***F(1,36)*=10,41 ; *p*<.01 ; η^2_p=.22**

Et on constate également un effet d'interaction entre ces deux facteurs :

***F(2,36)* = 4,06 ; *p*<.05 ; η^2_p=.18**

Figure 6-5 : Taux de prises de parole pendant le menu

On peut lire sur le graphique que les participants qui ont été confrontés aux énoncés de *'style condensé'* ont eu tendance à prendre la parole moins souvent pendant la présentation du menu par le système. Cependant, dans la condition de *'style espacé'* les participants qui ont eu une stratégie d'exploration directe prenaient la parole pendant le menu, comme ceux qui entendaient des *'énoncés normaux'*. Ainsi, la condition de *'style espacé'* rendait également les participants plus impatients.

- ***Taux de recouvrement entre parole de l'utilisateur et du système***

Le taux de recouvrement entre parole de l'utilisateur et parole du système est un indicateur important car il dépend de la structuration du dialogue par les participants. Les taux de recouvrement ont été comparés avec la même procédure GLM que les taux de prise de parole, ci-dessus (*'style d'énoncé'* et *'mode d'exploration'* en prédicteurs catégoriels). La Figure 6-6 présente le graphique obtenu pour la *'phase de détail'* à l'essai simple, comme cela a été présenté pour le moment de la prise de parole.

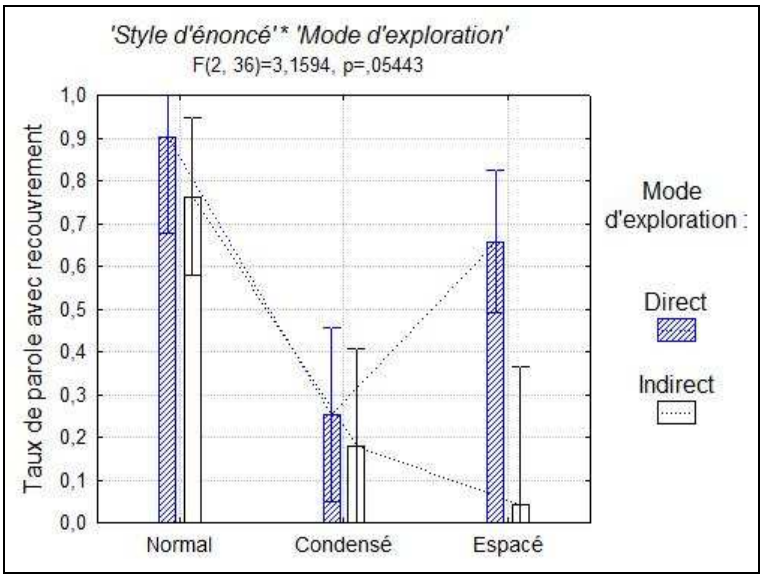


Figure 6-6 : Taux de prises de parole avec recouvrement

Le test de Levene révèle des distributions homogènes :
F(5,39)=0,55 ; *p*=.73

Le style d'énoncé a eu un effet sur le taux de recouvrement :
***F(2,36)*=16,6 ; *p*<.00 ; η^2_p=.48**

Le mode d'exploration utilisé a également eu un effet :
***F(1,36)*=9,35 ; *p*<.01 ; η^2_p=.20**

Et on constate également un **effet d'interaction marginal** entre le style d'énoncé et le mode d'exploration :
F(2,36)=3,15 ; *p*=.05 ; η^2_p=.14

On voit que la forme du graphique est identique à celle obtenue pour le taux de prises de parole pendant le menu. Ainsi, il semble que les utilisateurs qui ont pris la parole à ce moment ont provoqué un taux de recouvrement de leur parole avec celle du système plus important. Les énoncés de *'style condensé'* ont permis aux participants d'éviter un recouvrement fréquent de leur parole avec celle du système.

6.2.4 Conclusion

Globalement, les modifications de formulation des énoncés du système n'ont pas eu beaucoup d'effet sur les indicateurs généraux. Le rappel des informations n'a pas été affecté, pas plus que la charge cognitive subjective ni la performance globale dans la tâche. Par ailleurs, il n'y a pas eu plus d'erreurs de sélection des films dans une condition que dans l'autre. (Ce résultat n'a pas été présenté : le nombre de participants à avoir commis des

erreurs de sélection dans le scénario complexe était de trois pour le *'style normal'*, deux pour le *'style condensé'* et trois pour le *'style espacé'*.) On remarque seulement que la suppression de la syntaxe (styles *'condensé'* et *'espacé'*) a conduit les participants à écouter les informations sur les films un plus grand nombre de fois. Ce résultat peut être interprété selon deux points de vue distincts : (1) D'abord, une plus grande densité des informations conduit à un plus grand nombre d'écoutes, ce qui peut être vu, sous l'angle des ressources, comme l'effet de la limitation de capacité du canal auditif. Dans ce cas, c'est une augmentation de la charge cognitive auditive-verbale qui aurait imposé aux participants de réécouter les informations un plus grand nombre de fois pour obtenir la même performance de rappel. (2) D'un autre coté, ce résultat peut aussi être interprété, selon un point de vue comportemental, comme une plus grande facilité à réécouter les informations quand celles-ci sont présentées plus rapidement. Le confort d'écoute aurait alors été modifié indépendamment de l'effort et du rappel.

Du point de vue de la performance des utilisateurs dans la réalisation de la tâche, seul le nombre de mots a été légèrement affecté. Les participants ont fait plus de commandes en un seul mot quand le système produisait des énoncés de *'style espacé'*.

Du point de vue comportemental, les énoncés de *'style normal'* ont conduit les participants à prendre la parole le plus souvent *'pendant le menu'*. Ils se montraient ainsi plus réactifs, ou également plus impatients. Au contraire, les énoncés de 's*tyle condensé'* ont conduits les participants à prendre la parole plus tardivement. Dans ce sens, les participants confrontés au *'style condensé'* avaient une réactivité moindre, ou plus de patience. Quoiqu'il en soit, le taux de recouvrement entre leur parole et celle du système était moindre également. Dans le cadre du fonctionnement d'un système, des taux de recouvrement plus faibles peuvent permettre d'abaisser le taux d'hésitations de la part du participant et d'erreurs de reconnaissance vocale de la part du système.

Les utilisateurs se sont adaptés aux contraintes dans tous les cas. Ces résultats montrent qu'il est difficile de réduire la complexité des énoncés vocaux. L'utilisation complémentaire de la modalité visuelle pour présenter les énoncés du système pourrait permettre de produire des effets plus importants.

6.3 Conclusion des expériences 1 et 2

Les expériences 1 et 2 ont permis de décrire des dialogues humain-système selon des aspects à la fois cognitifs et comportementaux. L'expérience 1 a permis de constater que certaines séquences d'action rencontrées par les utilisateurs (les erreurs dès le premier dialogue) peuvent avoir une influence sur leur performance cognitive (le rappel) ; ce qui prouve, si c'est nécessaire, que ces deux aspects sont liés. Elle montre par ailleurs que l'économie de certaines informations (les *aides procédurales*) dans les énoncés du système

n'est pas souhaitable car l'utilisateur a besoin d'être correctement guidé. L'expérience 2 quant à elle, a permis de montrer que de légères modifications comportementales peuvent être obtenues si certaines règles de conception des énoncés sont appliquées, sans modifier nécessairement la performance dans la tâche ni la performance cognitive. On voit à travers ces expériences que le dialogue est une activité complexe car plusieurs processus sont à l'œuvre et multiplient les relations de causalité, qui deviennent difficiles à lister.

Pour cette raison, le processus dialogique est difficile à optimiser. Les énoncés vocaux sont consommateurs de temps et ils nécessitent un comportement d'écoute de la part de l'utilisateur. De plus, le comportement réactif des utilisateurs (les *prises de parole*) dépend du contenu et de la forme des énoncés. Dans l'expérience 1, les explications procédurales contenues dans les aides ont donné aux participants des connaissances qui les ont rendus plus réactifs dans leurs prises de parole. Dans l'expérience 2, c'est la forme des énoncés (forme linguistique : syntaxe et prosodie) qui a eu un effet sur la réactivité lors des prises de parole. Les expériences suivantes doivent permettre de vérifier si des effets de ce type, automatiques ou contrôlés, peuvent être obtenus grâce à la modalité visuelle.

Chapitre 7 Présentation audio-visuelle en DHM vocal

L'utilisation de la modalité visuelle en complément à la modalité auditive est un moyen qui pourrait permettre d'optimiser le processus dialogique. Les expériences 3, 4 et 5 ont pour objectif d'identifier des effets qui pourraient être obtenus par ce moyen. Mais d'abord, l'utilisation de la modalité visuelle implique une étape préalable de conception des énoncés audio-visuels, ce qui introduit la nécessité de catégoriser les contenus. En effet, si tous les contenus qui constituent les énoncés sont affiliés à une catégorie « information » unique, la seule possibilité qui s'offre au concepteur est un choix global entre présentation auditive, présentation visuelle et présentation redondante (bimodale) de l'ensemble de ces « informations ». La catégorisation des contenus en plusieurs éléments permet d'aller plus loin et de proposer ce choix pour chaque élément identifié, ce qui permet d'augmenter le nombre de possibilités. Plusieurs distinctions sont utilisables. Par exemple, Moreno et Mayer (1999) distinguent les informations graphiques des informations verbales. Il s'agit dans ce cas, d'une distinction liée aux propriétés des données en elles-mêmes. Dans les termes de Coutaz et Nigay (1994), cette distinction ne porte que sur le *'langage d'interaction'* qui permet d'encoder l'information, mais qui n'est pas suffisant pour définir une *'modalité'*.[1]. De même, il serait possible d'utiliser la distinction d'Anderson (1982) entre informations *procédurales* et *déclaratives*. Cette dichotomie est utilisée, par exemple, par Babin (2007) pour la conception de messages d'aide. Mais cette distinction ne permet pas de différencier les uns des autres les éléments à présenter dans un énoncé. Dans le cadre de l'étude du processus dialogique, il est également possible d'utiliser une catégorisation fonctionnelle, liée à l'activité de dialogue en elle-même. C'est le cas de la distinction entre : *'écho'*, *'réponse'* et *'relance'* (tirée des travaux de Nievergelt et Weydert, 1980) qui a été introduite au chapitre 2. Que les informations soient verbales ou graphiques, procédurales ou déclaratives, elles peuvent être affiliées à l'une ou l'autre de ces trois catégories en fonction du (ou des) rôle(s) qu'elles jouent dans l'activité. L'avantage de cette catégorisation est de permettre l'attribution d'un rôle à chaque élément et d'être en même temps indépendante du contexte, ce qui permet de généraliser la problématique. Ainsi, les énoncés testés dans les expériences 3, 4 et 5 ont été conçus sur la base de cette distinction de façon à étudier le processus de dialogue dans sa perspective fonctionnelle.

Les questionnements à la base des trois expériences qui suivent sont liés aux effets des différents types d'information dans les dialogues quand le mode de présentation change. L'objectif est de mettre en évidence, autant que possible, la diversité des effets produits de

[1] D'après la définition de Coutaz et Nigay (1994) une « modalité » correspond à un couple (*'dispositif technique'*, *'langage d'interaction'*).

façon à identifier une (ou des) combinaison(s) qui conduirai(en)t à une plus grande utilisabilité (efficacité, efficience, satisfaction, *Cf.* ISO 9241-11) des services de DHM. L'expérience 3 était focalisée sur l'impact en mémoire des modes de présentation et sur le gain supposé qui pourrait être associé à une présentation bimodale. Les expériences 4 et 5 ont permis d'étudier les dialogues de façon plus complète. Elles permettent de mettre en évidence la relation spécifique entre modalité perceptive et type d'information dans le processus dialogique.

7.1 Expérience 3 : Redondance audio-visuelle et effet de suffixe

7.1.1 Objectifs et hypothèses

L'expérience 3 a été conçue pour vérifier si la présentation audio-visuelle des informations dans le contexte du DHM pouvait permettre d'éviter l'*effet de suffixe*, identifié dans le cadre des travaux sur la mémoire de travail (voir Penney, 1989, pour une revue).

L'effet de suffixe a été identifié dans des tâches de mémorisation de listes d'items. Il correspond à l'effet négatif d'un mot inutile placé en fin de liste, qui fait baisser la performance au test de rappel (voir Crowder, 1967; Engle, 1974; Hitch, 1975). Du fait de la présence de ce mot, qui ne doit pas être mémorisé, après la fin de la liste, les mots à mémoriser de la fin de la liste sont oubliés plus fréquemment. Cet effet est moins important si le suffixe n'a pas de contenu verbal (simple « bip »). D'autre part, il est plus important avec la modalité auditive qu'avec la modalité visuelle. Par exemple, dans une expérience testant plusieurs types de suffixes, Hitch (1975) est parvenu à éliminer l'effet de suffixe dans certaines conditions si la liste était présentée visuellement, mais pas si elle était présentée auditivement. Quand la liste était présentée auditivement, tous les suffixes produisaient un effet négatif. Ce résultat indique que l'effet de suffixe pourrait jouer un rôle important dans le dialogue puisque celui-ci repose sur la présentation auditive des informations. Mais d'un autre coté, dans une étude réalisée avec du matériel linguistiquement cohérent (définitions du dictionnaire et histoires courtes) Balota, Engle et Cowan (1990) ont montré que l'effet de suffixe était moins important si des connaissances antérieures telles que les connaissances linguistiques permettent de mieux structurer l'information. Ainsi, l'effet de suffixe observé dans des conditions expérimentales « dé-contextualisées » pourrait jouer un rôle de plus faible importance en contexte. Mais cette hypothèse doit être vérifiée dans un contexte de DHM.

La situation expérimentale correspondant à l'effet de suffixe peut facilement être appliquée au cas de la présentation des énoncés d'un système de dialogue. Les *'réponses'* du système se présentent sous la forme d'une liste d'éléments que l'utilisateur doit comprendre et mémoriser. Les *'relances'* se présentent sous la forme d'une phrase d'ouverture, qui peut agir comme un suffixe. Cet effet est d'autant plus probable que les *'relances'* présentes dans le DHM sont linguistiquement plus riches que les suffixes en un seul mot généralement utilisés dans les expériences (*e.g.* Hitch, 1975). Ainsi, l'expérience 3 doit permettre de *mesurer l'effet d'interférence des 'relances' sur la mémorisation des 'réponses'* dans le contexte du DHM.

On peut aussi se demander, suivant l'hypothèse additive, si la présentation bimodale (redondante) des *'réponses'* facilite la mémorisation et/ou si elle peut permettre de prévenir l'effet de suffixe. D'un autre coté, la présentation bimodale pourrait surcharger l'utilisateur, ce qui doit également être vérifié.

Par ailleurs, le point de vue pragmatique permet de supposer que la présentation audio-visuelle des informations produit des effets spécifiques dans le processus interactif. Les participants adapteront leurs actions aux possibilités offertes. Aucune hypothèse spécifique n'était formulée à ce sujet lors de la conception de l'expérience.

7.1.2 Méthode

A Participants

Le nombre final de participants était de 54 (10 hommes, 44 femmes). La moyenne d'âge du groupe était de 20,81 ans (*é.t.* = 1,40 ; *min.* = 19 ; *max.* = 26). Les participants étaient des étudiants en psychologie à l'Université Rennes 2 en Licence – première année. Ils participaient à l'expérience dans le cadre des enseignements en psychologie expérimentale, en tant qu'activité obligatoire associée au cours.

Le niveau d'expérience en informatique des participants était moyen. Ils ont déclaré pratiquer l'informatique depuis 6,18 ans en moyenne (*é.t.* = 3,48 ; *min.* = 0 ; *max.* = 15). 17 d'entre eux ont déclaré disposer d'un ordinateur régulièrement. Les 37 autres participants ne connaissaient que les applications bureautiques.

En ce qui concerne la pratique des services téléphoniques, 5 participants ont déclaré utiliser régulièrement des services vocaux interactifs, 21 occasionnellement, 17 rarement et 10 jamais. Les services mentionnés étaient parmi les plus courants (répondeur, renseignements). La pratique des services à reconnaissance vocale était peu répandue dans cette population.

B Matériel

Le système utilisé pour cette expérience était une simulation du service *'PlanResto'*. (*Cf.* Figure 5-2) développée et mise en place selon le dispositif expérimental présenté au chapitre 5. Ce service correspond à un dialogue en trois phases (*'Requête'–'Sélection'–Information'*) dédiées au choix d'un restaurant. La structure du dialogue avec ce service imposait de répéter plusieurs fois la phase de *'Sélection'* car les restaurants étaient présentés un par un. C'est pour cette raison que ce service n'a pas été utilisé dans les autres expériences de la thèse, car cette structure limite les initiatives des participants. Dans le cas de l'expérience 3, l'objectif était de limiter les initiatives des participants pour les amener à la situation voulue de présentation des informations sur un restaurant suivie d'une *'relance'*. La vérification des hypothèses concernant l'*effet de suffixe* et l'interférence éventuelle des *'relances'* reposait sur cette mise en situation. Ce service était intéressant pour répondre à ce besoin.

C Procédure

Le participant était accueilli individuellement. Il était conduit dans une pièce calme où était installé le matériel expérimental. L'expérimentateur lui présentait brièvement, à l'oral, le

service *'PlanResto'* et le matériel qu'il s'apprêtait à utiliser pour réaliser les appels. Il lui remettait ensuite une feuille de papier présentant un tableau de 12 cellules numérotées pour le rappel des informations sur les restaurants consultés au cours des 12 scénarios expérimentaux. Comme dans l'expérience 2, la présentation de cette feuille impliquait que la *consigne de rappel devenait explicite*. De ce fait, pour limiter l'impact des adaptations comportementales dans le dialogue, l'expérimentateur expliquait au participant qu'une seule consultation des informations sur le restaurant choisi était permise. La consigne indiquait au participant de choisir un restaurant susceptible de lui plaire, d'en consulter les informations et de demander la mise en relation comme s'il souhaitait faire une réservation.

Après avoir pris connaissance de la procédure expérimentale, le participant commençait l'expérience sur l'ordinateur dédié. La procédure était organisée par le programme de test. Celui-ci incluait une page de présentation du questionnaire d'évaluation subjective de la charge de travail (questionnaire NASA-TLX), qui était affichée avant le premier dialogue. Trois évaluations étaient ensuite demandées au cours de l'expérience ; (1) après le quatrième dialogue, (2) après le huitième et (3) après le douzième, c'est-à-dire à chaque changement de mode de présentation des *'réponses'* (facteur 1, voir ci-dessous).

Après avoir réalisé l'ensemble des scénarios expérimentaux, le participant restait pour un entretien de quelques minutes au cours duquel il répondait à l'oral aux questions de l'expérimentateur sur son expérience des services téléphoniques, son expérience en informatique et ses sentiments concernant le service qu'il venait d'utiliser.

D Protocole expérimental

Deux variables indépendantes étaient manipulées dans cette expérience.

- ***Facteur 1* : Mode de présentation des « réponses »**[1]

Trois modes de présentation ont été utilisés pour présenter les *'réponses'* du système.

1. ***'Visuel'*** : Les *'réponses'* étaient affichées à l'écran ;
2. ***'Redondant'*** : Les *'réponses'* étaient affichées à l'écran et diffusées vocalement ;
3. ***'Auditif'*** : Les *'réponses'* étaient diffusées vocalement.

Il s'agit d'une *variable intragroupe*. Chaque participant faisait 4 dialogues successifs dans chaque condition, soit 12 dialogues au total. La position des conditions était contre-balancée et constituait 6 sous-groupes différents correspondant à l'arrangement des trois conditions.

[1] **NB** : Le terme *'réponse'* n'est pas utilisé ici dans son sens général. Il désigne la catégorie de contenu présentée par le système qui correspond aux informations recherchées par l'utilisateur (voir chapitre 2). Dans la présentation des résultats, comme dans l'ensemble du document, ce terme est toujours présenté entre des cotes et en italique (*'réponse'*) et ne doit pas être confondu avec les « temps de réponse » des participants, utilisés comme indicateurs comportementaux.

- ***Facteur 2 : Mode de présentation des « relances »***

Trois modes de présentation ont été utilisés pour présenter les *'relances'* du système.

1. ***'Visuel'*** : Les *'relances'* étaient affichées à l'écran ;
2. ***'Redondant'*** : Les *'relances'* étaient affichées à l'écran et diffusées vocalement ;
3. ***'Auditif'*** : . Les *'relances'* étaient diffusées vocalement.

Il s'agit d'une *variable intergroupe*. Chaque participant n'a été confronté qu'à un seul type de relance au cours des essais expérimentaux.

E Mesures dépendantes

Les mesures utilisées dans la présentation des résultats sont indiquées dans le Tableau 7-1. Lorsque des explications sur la signification des différents indicateurs sont nécessaires, elles sont précisées lors de la présentation des résultats.

Tableau 7-1 : Indicateurs utilisés pour la présentation des résultats de l'expérience 3

Rappel	Nombre d'items rappelés parmi 9. Les items à rappeler étaient issus des éléments suivant : *'nom du restaurant'*, *'arrondissement'*, *'adresse'* et *'numéro de téléphone'*).
Charge de travail subjective	Evaluée avec le questionnaire *'NASA-TLX'*.
Durée des dialogues	Durée totale (en secondes)
Temps de réponse suite aux relances	Les temps de réponse ont été mesurés (Figure 5-2) : - Dans la *« phase de sélection »* et ; - Dans la *« phase d'information »*. Ces temps étaient décomptés à partir de la fin de l'énoncé du système, *i.e.* à la fin de la relance dans les conditions *'relance auditive'* et *'relance redondante'*. et après le temps nécessaire à la lecture de la relance dans la condition *'relance visuelle'*.

7.1.3 Résultats

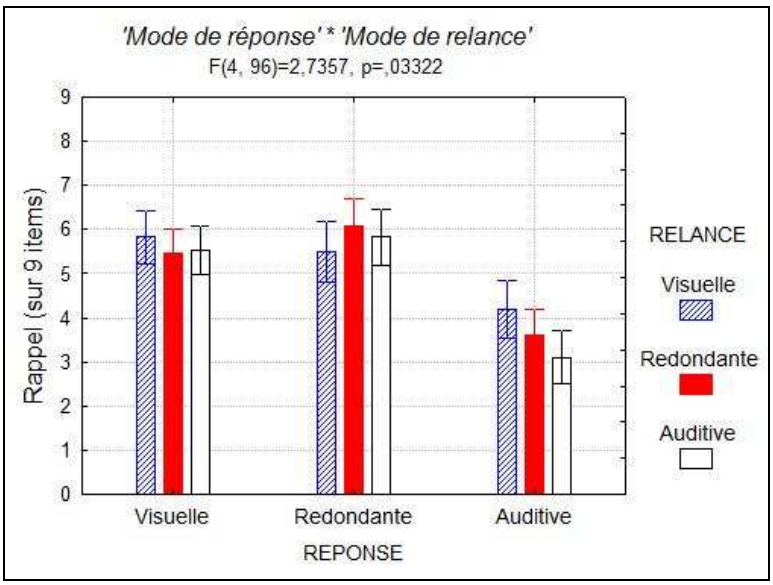
Le corpus final est composé de 648 dialogues. Ces dialogues n'ont pas été transcrits dans la mesure où cette expérience était focalisée sur la mémorisation des *'réponses'* du système. Les résultats présentés sont la performance au test de rappel, l'évaluation subjective de la charge de travail (avec le questionnaire NASA-TLX), la durée des dialogues et le temps de réponse aux *'relances'* dans la *'phase de sélection'* et dans la *'phase d'information'* du dialogue avec le service *PlanResto* (*Cf.* Figure 5-2).

Tous les résultats présentés sont des indicateurs numériques continus. Ils sont analysés avec des procédures GLM utilisant le mode de présentation des *'relances'* comme prédicteur catégoriel et le mode de présentation des *'réponses'* en tant que mesure répétée.

A Indicateurs cognitifs généraux

- ***Rappel des informations sur le restaurant choisi***

Les scores de rappel considérés sont les moyennes des scores de rappel obtenus aux quatre essais successifs dans chaque condition. Un score moyen était obtenu pour chaque mode de *'réponse'*. Les tests de Levene ont indiqué que ces scores avaient des variances homogènes (*F(2,51)*=0,25 ; *p*=.77 pour les *'réponses'* visuelles ; *F(2,51)*=0,47 ; *p*=.62 pour les *'réponses'* redondantes ; et *F(2,51)*=1,08 ; *p*=.34 pour les *'réponses'* auditives). Pour la comparaison, le temps de réponse en *'phase d'information'* (présenté plus loin) a été utilisé en co-variable.



Le mode de présentation des *'relances'* n'a pas eu d'effet significatif sur le score de rappel :

F(2,48)=0,50 ; *p*=.60 ; η^2_p=.02

Le mode de présentation des *'réponses'* a produit un effet significatif :

***F(2,96)*=10,34 ; *p*<.00 ; η^2_p=.17**

Et il y avait un effet d'interaction entre ces deux facteurs :

***F(4,96)*=2,73 ; *p*<.05 ; η^2_p=.10**

Figure 7-1 : Rappel des informations sur le restaurant

On voit sur le graphique que les *'réponses'* auditives ont donné lieu à des scores de rappel plus faibles que les *'réponses'* visuelles et redondantes. L'effet d'interaction est lié à la dégradation des scores de rappel dans la condition où *'réponse'* et *'relance'* sont toutes les deux présentées auditivement. Ainsi, l'*effet d'interférence des relances auditives* n'est apparu que lorsque toutes les informations étaient présentées auditivement.

- ***Charge de travail subjective***

Dans cette expérience, la charge de travail a été évaluée avec le questionnaire NASA-TLX. L'échelle en 15 points utilisée pour calculer l'indice correspond à la version initiale du questionnaire (*Cf.* Hart & Staveland, 1988). Les tests de Levene ont révélés des données homogènes pour les *'réponses'* visuelles (*F(2,51)*=0,50 ; *p*=.60), pour les *'réponses'* redondantes (*F(2,51)*=1,18 ; *p*=.31) et pour les *'réponses'* auditives (*F(2,51)*=0,62 ; *p*=.53).

On voit sur le graphique que les participants ont évalué la charge de travail comme étant plus importante quand les *'réponses'* étaient présentées auditivement. Ils ont trouvé cette condition plus difficile que les deux autres. Autrement dit, l'effort nécessaire à l'accomplissement du dialogue a été sensible au mode de présentation des *'réponses'* – les *'réponses'* visuelles ont nécessité un effort moins important – mais pas au mode de présentation des *'relances'* – qui était indifférent, de ce point de vue.

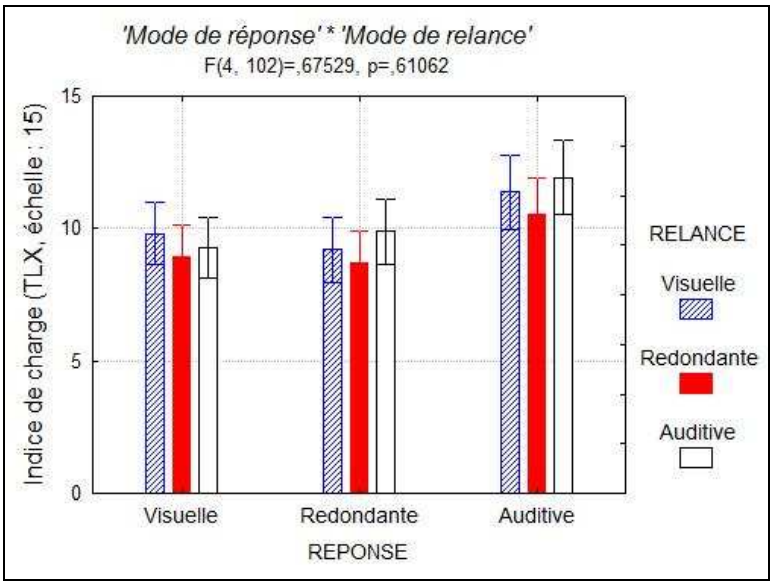


Figure 7-2 : Indice de charge de travail subjective (TLX)

Aucune co-variable n'a été utilisée dans la procédure GLM.

Le mode de présentation des *'relances'* n'a pas eu d'effet :

F(2,51)=0,98 ; *p*=.38 ; η^2_p=.03

Le mode de présentation des *'réponses'* a produit un effet très significatif sur les indices TLX :

***F(2,102)*=19,7 ; *p*<.00 ; η^2_p=.27**

Et il n'y a pas eu d'effet d'interaction entre ces deux facteurs :

F(4,102)=0,67 ; *p*=.61 ; η^2_p=.02

B Indicateur de réalisation de la tâche

Dans la mesure où la transcription des dialogues n'est pas disponible, le seul indicateur de réalisation de la tâche proposé ici est la durée des dialogues.

Les tests de Levene ont révélé des données homogènes (*F(2,51)*=0,95 ; *p*=.39 pour les *'réponses'* visuelles ; *F(2,51)*=0,73 ; *p*=.48 pour les *'réponses'* redondantes ; et *F(2,51)*=0,67 ; *p*=.51 pour les *'réponses'* auditives).

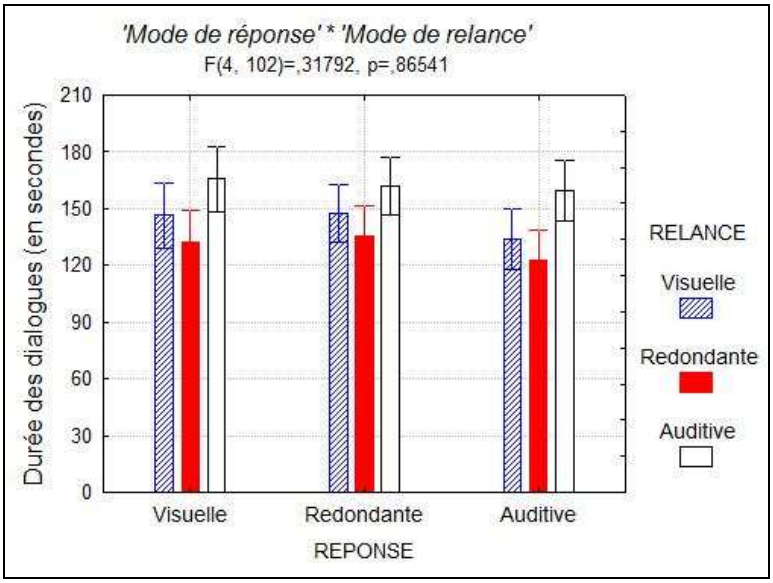


Figure 7-3 : Durée des dialogues

Le mode de présentation des *'relances'* a eu un effet sur la durée des dialogues :

***F(2,51)*=6,44 ; *p*<.01 ; η^2_p=.20**

Le mode de présentation des *'réponses'* a produit un effet **marginalement significatif** :

F(2,102)=2,63 ; *p*=.07 ; η^2_p=.04

Et il n'y a pas eu d'effet d'interaction entre ces deux facteurs :

F(4,102)=0,31 ; *p*=.86 ; η^2_p=.01

On lit sur le graphique de la Figure 7-3 que les *'relances'* redondantes ont donné lieu aux dialogues les plus courts. Les dialogues les plus longs ont été obtenus pour les *'relances'* auditives, résultat qui ne correspond pas aux temps de réponse présentés plus loin. Il indique en fait que dans la *'phase de requête'*, en début de dialogue, les participants ont pris la parole moins tôt quand les *'relances'* étaient auditives, car ils attendaient que l'énoncé du système

soit complet. Ce résultat est peu significatif vis-à-vis du test de rappel de l'expérience. Mais on peut cependant remarquer que la forme du graphique obtenue pour la durée des dialogues (Figure 7-3) est proche de celle obtenue pour l'indice d'effort (Figure 7-2), ce qui suggère le lien, connu, entre durée et coût de l'activité.

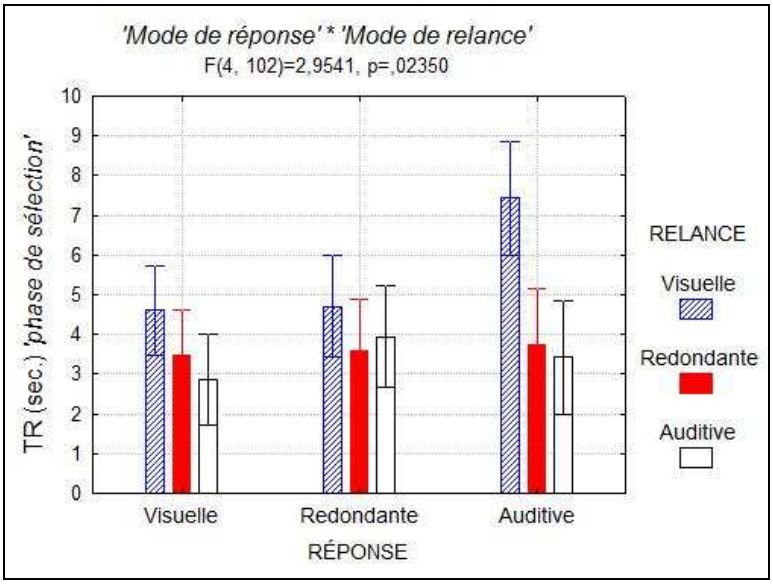
C Indicateurs comportementaux

Les deux indicateurs de comportement stratégique proposés ici sont les temps de réponse aux *'relances'* (1) dans la *'phase de sélection'*, dans laquelle le système propose des informations générales sur le restaurant, et (2) dans la *'phase d'information'*, dans laquelle le système propose le détail des informations. Le participant devait simplement faire un choix dans le premier cas, alors qu'il devait mémoriser la *'réponse'* du système dans le second. C'est pourquoi les temps de réponse sont plus longs dans le second cas.

Dans chaque dialogue le participant (1) formulait une requête, (2) consultait les informations générale sur le premier restaurant, puis sur les suivants jusqu'à ce qu'il identifie un restaurant susceptible de lui plaire, (3) il demandait le détail des informations sur ce restaurant, et enfin (4) il demandait la mise en relation, ce qui terminait l'appel. Les temps de réponse présentés ci-dessous correspondent aux temps enregistrés aux étapes (3) et (4) de cette séquence. Comme pour le rappel, les temps de réponse présentés sont des moyennes des quatre essais réalisés dans chaque condition.

- ***Temps de réponse en 'phase de sélection'***

Pour la *'phase de sélection'*, les tests de Levene ont révélé des données homogènes pour les temps de réponse obtenus pour les *'réponses'* visuelles ($F(2,51)$=0,41 ; p=.66) et pour les *'réponses'* redondantes ($F(2,51)$=0,64 ; p=.52), mais pas pour les *'réponses'* auditives ($F(2,51)$=5,20 ; p<.01). La procédure GLM pour comparer les moyennes a été appliquée malgré la significativité du dernier de ces trois résultats.



Le mode de présentation des *'relances'* a eu un effet sur les temps de réponse dans cette phase :

***F(2,51)*=6,72 ; *p*<.01 ; η^2_p=.20**

Le mode de présentation des *'réponses'* a aussi produit un effet significatif :

***F(2,102)*=3,79 ; *p*<.05 ; η^2_p=.07**

Et il y avait un effet d'interaction entre ces deux facteurs :

***F(4,102)*=2,95 ; *p*<.05 ; η^2_p=.10**

Figure 7-4 : Temps de réponse en *'phase de sélection'*

On lit sur le graphique de la Figure 7-4 que les temps de réponse des participants étaient plus longs pour les *'relances'* visuelles que pour les deux autres modes de présentation. Cette différence était particulièrement évidente quand les *'réponses'* étaient présentées auditivement puisque dans cette condition il doublait (de 4 secondes à 8 secondes). Ainsi, quand la *'réponse'* était auditive et la *'relance'* visuelle les participants ont pris un temps de réflexion plus important avant de faire leur choix.

- ***Temps de réponse en 'phase d'information'***

De même que pour la *'phase de sélection'*, les tests de Levene pour les temps de réponse en *'phase d'information'* ont révélé des données homogènes pour les *'réponses visuelles'* ($F(2,51)$=0,59 ; p=.55) et pour les *'réponses redondantes'* ($F(2,51)$=0,36 ; p=.69), mais pas pour les *'réponses auditives'* ($F(2,51)$=3,39 ; p<.05). De même, la comparaison de moyenne a été effectuée malgré la significativité du troisième de ces résultats.

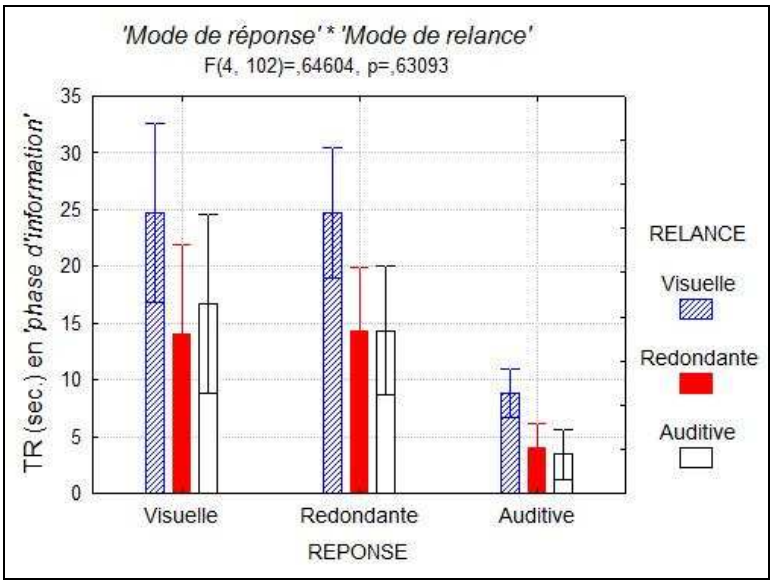


Figure 7-5 : Temps de réponse en *'phase d'information'*

De même, le mode de présentation des *'relances'* a eu un effet sur les temps de réponse dans cette phase :

$F(2,51)$=4,82 ; p<.02 ; η^2_p=.16

Le mode de présentation des *'réponses'* a aussi produit un effet significatif :

$F(2,102)$=31,6 ; p<.00 ; η^2_p=.38

En revanche, il n'y avait pas d'effet d'interaction entre les deux facteurs :

$F(4,102)$=0,64 ; p=.63 ; η^2_p=.02

On lit sur le graphique de la Figure 7-5 que les temps de réponse étaient de 5 secondes pour une *'réponse'* auditive et une *'relance'* soit auditive soit bimodale. Ce temps passait à près de 10 secondes quand la *'relance'* n'était présentée que visuellement. Dès que la *'réponse'* était présentée visuellement, les temps de réponse des participants passaient à 15 secondes après les *'relances'* auditives ou bimodales et à 25 secondes pour les *'relances'* visuelles. Deux effets se sont conjugués : (1) la présentation visuelle des *'réponses'* (mode visuel et mode redondant) a permis aux participants de prendre le temps de relire les informations avant de demander la mise en relation et (2) la présentation auditive des *'relances'* (modes auditif et redondant) les a incités à demander plus rapidement la mise en relation.

7.1.4 Conclusion

La supériorité du rappel pour les *'réponses'* écrites (conditions visuelle et redondante) est liée à la possibilité de relire les informations présentées à l'écrit (tant que la commande suivante

n'est pas prise en compte). Dans ces conditions, les participants ont exploité cette possibilité. Ils ont respecté la consigne qui leur était donnée de ne consulter qu'une seule fois les informations, mais ils les ont consulté plus longuement, ce qui leur a permis d'obtenir un meilleur score de rappel.

L'utilisation de cette consigne était liée à un souci d'équité entre les modalités, en vue du test de rappel. Il s'agissait de limiter le comportement adaptatif des participants (la demande active de « réécouter ») pour *donner autant de chance* – si l'on peut dire – à chacun des canaux perceptifs. Or, ce n'est pas le résultat qui a été obtenu puisque la modalité visuelle a permis aux participants de relire les informations avant de demander la mise en relation. C'est un autre comportement adaptatif qui a pris le pas sur celui qui était éliminé. Il semble donc que l'utilisation de cette consigne n'ait pas été très utile. Il serait peut-être plus judicieux d'admettre que, dans le dialogue, les utilisateurs ont des degrés de liberté qui leur permettent d'atteindre leurs objectifs comme ils l'entendent. Pour étudier ce processus, il est peut-être préférable de ne pas chercher à le contraindre pour l'adapter à la théorie, mais au contraire, d'y laisser libre cours afin de tenter de comprendre comment y adapter la théorie. Pour cette raison, ce type de contrainte n'a pas été imposé aux participants des autres expériences.

Malgré cette restriction, les résultats obtenus présentent de l'intérêt. Un effet d'interférence des *'relances'* auditives a bien été constaté sur la mémorisation des *'réponses'* auditives, ce qui indique que l'effet de suffixe joue bien un rôle dans le DHM vocal.

Autour de cet effet de suffixe, on peut distinguer l'effet du mode de *'réponse'* de l'effet du mode de *'relance'*. Les résultats obtenus pour les *'réponses'* visuelles et redondantes offrent la solution au problème. Ils indiquent que *la présentation visuelle de ce type d'information permet de prévenir l'effet de suffixe*. Cet effet bénéfique est spécifique au mode de présentation des *'réponses'*. Par ailleurs, l'effet de suffixe est l'effet négatif des *'relances'* auditives. Il n'a été observé que pour les *'réponses'* auditives, dont la relecture était impossible. Mais les *'relances'* auditives ont joué un rôle également quand les *'réponses'* étaient présentées visuellement : les temps de réponse étaient plus courts. Ainsi, les *'relances'* auditives semblent permettre de dynamiser le dialogue. Elles jouent d'autres rôles que celui qui consiste à provoquer l'effet de suffixe. Les questions posées dans les *'relances'* étaient, pour la *'phase de sélection'* : *« Voulez-vous plus d'informations, les restaurants suivants ou effectuer une nouvelle recherche ? »* et pour la *'phase d'information'* : *« Souhaitez-vous être mis en relation avec ce restaurant, les restaurants suivants ou effectuer une nouvelle recherche ? »* Ces questions évoquent les possibilités d'action à suivre, ce qui focalise l'attention des utilisateurs sur cet aspect. Ainsi, les temps de réponses plus courts après la présentation auditive de ces informations met en évidence un effet spécifique au mode de présentation des *'relances'* (focalisation sur l'objectif). Il semble donc qu'il y ait une relation particulière entre le *'type d'information'* à présenter (et le *type* renvoie au *rôle* de l'élément au sein de l'*activité*) et le *'mode de présentation'* utilisé pour présenter *cette* information. Les expériences 4 et 5 ont pour objectif d'aller un peu plus loin dans l'étude de cette relation.

Par ailleurs, ces résultats mettent en évidence les liens entre l'activité déployée par les participants au cours de la réalisation de la tâche et la performance cognitive obtenue. D'abord, l'élévation de la charge de travail subjective pour les *'réponses'* auditives, ainsi que le raccourcissement des temps de réponses dans cette condition, peuvent être rapprochés de l'impossibilité de relire les informations présentées auditivement. Cette limite comportementale a pu accroître la difficulté de la tâche en imposant aux participants une activité mentale d'autorépétition (boucle phonologique de Baddeley). Cette activité mentale supplémentaire a pu conduire les participants à évaluer la tâche comme étant plus difficile que dans les autres conditions. Ensuite, on constate que les comportements des participants ont pris des formes différentes quand les modes de présentation des différents *types d'informations* changeaient. Les temps de réponse dans les phases de sélection (Figure 7-4) et d'information (Figure 7-5) sont éloquents à ce sujet. La présentation visuelle des *'réponses'* et la présentation auditive des *'relances'* ont eu des effets clairement distincts dans la *'phase d'information'*. Les *'réponses'* visuelles étaient consultées plus longtemps (effet spécifique au mode de *'réponse'*). Les *'relances'* auditives, quant à elles, étaient mieux détectées et faisaient réagir plus vite les participants (effet spécifique au mode de *'relance'*). Les participants ont hérité (en bénéfice ou en perte) des propriétés des modes de présentation dès qu'elles apparaissaient dans le courant de l'interaction : les *'réponses'* redondantes ont produit l'effet des *'réponses'* visuelles ; les *'relances'* redondantes on produit l'effet des *'relances'* auditives. Ces effets sont comportementaux (temps de réponse) et ils peuvent être rapprochés des effets sur les indicateurs cognitifs généraux (rappel et charge). La présentation visuelle des *'réponses'* a permis un meilleur rappel et a fait baisser la charge de travail subjective parce que les participants ont pris le temps de relire les informations avant de demander la mise en relation. La présentation auditive des *'relances'* a provoqué un effet de suffixe parce que les participants ont fait plus attention aux questions du système (réponse plus rapide). L'effet de suffixe semble donc également avoir des bases comportementales.

Il semble ainsi que les effets observés concernant les indicateurs cognitifs généraux trouvent leurs fondements dans les adaptations comportementales des participants plutôt que, plus strictement, dans les *capacités perceptives* de leurs *canaux sensoriels*. Pris sous cet angle, ces résultats tendent à favoriser une perspective pragmatique qui s'intéresse au déploiement de l'activité des individus, plutôt qu'une perspective cognitive, plus focalisée sur le codage de l'information.

7.2 Expérience 4 : Mise en évidence de la spécificité modale

7.2.1 Objectifs et hypothèses

Les résultats de l'expérience 3 ont montré que le *mode de présentation* des différents *types d'informations* produit des effets différents d'un *type d'information* à l'autre. Par ailleurs, le mode de présentation redondant (ou bimodal) n'a pas produit d'effet spécifique. Il s'apparentait soit au mode visuel, soit au mode auditif selon les cas. En conséquence, l'expérience 4 a été conçue pour mettre en évidence la relation entre type d'information et mode de présentation (ou *« relation type-mode »*). Comme on l'a vu, cette relation ne semble par reposer strictement sur les *capacités perceptives* des utilisateurs, mais plutôt sur les *comportements adaptatifs* qu'ils sont en mesure de déployer face à différents énoncés. C'est pourquoi, comme cela est présenté ci-dessous, cette relation a été étudiée sur la base d'une dissociation fonctionnelle (catégorisation de Bunt).

La démarche a consisté (1) à considérer l'ensemble des informations présentées dans les énoncés d'un système de DHM ; (2) à les catégoriser en leurs éléments (fonctionnels) ; (3) à établir une règle d'attribution des modes de présentation à ces éléments, ce qui a permis de définir les *« stratégies de présentation »* des énoncés et ; (4) à évaluer les effets de ces *« stratégies de présentation »* dans un contexte de DHM. L'objectif de cette démarche était d'identifier la stratégie de présentation la plus adaptée au DHM, selon une perspective d'optimisation des systèmes de dialogue.

La notion de *« stratégie de présentation »* est utilisée car l'attribution des modes de présentation aux différents types d'informations est un choix qui peut provoquer des effets divers dans le dialogue et dont l'impact peut varier en fonction de certains éléments contextuels. Il est probablement possible de prédéfinir des *« stratégies de présentation »* plus efficaces dans la plupart des cas. Mais il est également probable que de telles stratégies ne seront pas adaptées à tous les cas. Ainsi, le choix d'une *« stratégie de présentation »* à un moment donné dans un dialogue donné pourrait nécessiter un processus inférentiel (un raisonnement) plus complexe que la simple application de formes prédéfinies. Quoiqu'il en soit, la notion de *« stratégie »* est utilisée pour évoquer l'aspect évolutif, coordonné et orienté du processus (le dialogue) au sein duquel s'insère l'énoncé du système.

La règle d'attribution utilisée pour affecter les *modes de présentation* aux *types d'information* repose sur la catégorisation des fonctions de contrôle du dialogue proposée par Bunt (1994, *voir* chapitres 2). Ces fonctions renvoient aux obligations des partenaires d'une communication (s'ils souhaitent réussir leur communication, voir Schegloff, 2006). D'après le modèle pragmatique, chaque acte produit dans la conversation est destiné à remplir une ou plusieurs de ces fonctions. Au premier niveau de cette catégorisation, on distingue les *« fonctions de contrôle »* du dialogue des *« fonctions d'information »* (voir Figure 2-4, p. 55).

Ce premier niveau était suffisant pour la conception du matériel expérimental. L'une des deux fonctions a été attribuée, exclusivement de l'autre (pour les besoins de l'expérience), à chaque type d'information (*'écho'*, *'réponse'* et *'relance'*) de façon à vérifier si les obligations associées à chaque type sont mieux assurées par un mode de présentation que par l'autre :

- Les *'échos'* et les *'relances'* servent à vérifier les actions antérieures et à connaître les actions futures possibles. C'est la *fonction de contrôle* qui leur a été attribuée.
- Les *'réponses'* sont les informations cible du dialogue. C'est la *fonction d'information* qui leur a été attribuée.

La conception des *'stratégies de présentation'* a consisté à apparier les deux fonctions principales dans l'activité de dialogue (contrôle *vs.* information) aux deux modes de présentation disponibles (auditif *vs.* visuel). Cet appariement donne lieu à quatre combinaisons, soit quatre *« stratégies de présentations des informations »* :

(1) Contrôle Auditif – Information Auditive ;

(2) Contrôle Auditif – Information Visuelle ;

(3) Contrôle Visuel – Information Auditive et ;

(4) Contrôle Visuel – Information Visuelle.

Les prédictions de résultat étaient basées sur la distinction fonctionnelle utilisée (Bunt, 1994). Chaque fonction implique des sous-buts différents. L'information contenue dans les *'réponses'* doit être comprise, comparée, mémorisée, ré-accédée, etc. Nous avons supposé que la *modalité visuelle* offrirait des propriétés plus adaptées pour cette *fonction d'information*, par rapport à la modalité auditive. Inversement, l'information contenue dans les *'échos'* et les *'relances'* doit être prise en compte immédiatement dans l'interaction, par exemple, pour corriger une erreur de reconnaissance vocale ou donner des informations manquantes. Nous avons supposé que la *modalité auditive* offrirait des propriétés plus adaptées pour cette *fonction de contrôle*, par rapport à la modalité visuelle. C'est donc la seconde stratégie (*contrôle auditif – information visuelle*) qui était supposée supporter les meilleures performances. Ces prédictions fonctionnelles relèvent de l'*hypothèse pragmatique*.

L'*hypothèse cognitive*, basée sur l'*additivité* des ressources cognitives, amène à une prédiction différente. Les deux stratégies bimodales (2) et (3) sont supposées accroître la capacité de mémoire de travail disponible et permettre un meilleur rappel. En effet, les stratégies proposées ne sont pas redondantes (même information présentée plusieurs fois) et elles ne correspondent pas aux cas de dissociation de l'attention identifiés dans la littérature. Les deux stratégies bimodales sont conçues suivant le principe d'utilisation de modalités multiples, supposé améliorer l'apprentissage pour des novices (Kalyuga, Chandler & Sweller, 1998; Mousavi, Low & Sweller, 1995). Par ailleurs, l'effet de suffixe disqualifie la stratégie (1) (entièrement auditive) qui devrait donner lieu au rappel le plus faible ; et la préférence pour la présentation visuelle des *'réponses'* privilégie les stratégies (2) et (4) qui devraient favoriser le rappel.

7.2.2 Méthode

A Participants

Le nombre final de participants était de 80 (10 hommes, 70 femmes). La moyenne d'âge du groupe était de 19 ans (*é.t.* = 1,76 ; *min.* = 17 ; *max.* = 26). Les participants étaient des étudiants de l'université Rennes 2 en provenance de plusieurs disciplines et s'étalant de la première année d'étude après le bac au Master 2. Chaque participant a reçu un bon d'achat d'une valeur de 10 € pour sa participation.

Le niveau de pratique de l'informatique des participants était plus élevé que dans les autres expériences d'après leurs déclarations. Ils ont cité en moyenne 7,06 logiciels qu'ils déclarent utiliser régulièrement (*é.t.* = 3,27 ; *min.* = 2 ; *max.* = 16) et ils ont déclaré passer 22,95 heures par mois sur internet (*é.t.* = 11,22 ; *min.* = 1 ; *max.* = 100), soit trois fois plus que les participants aux expériences 1 et 2. En ce qui concerne les services vocaux téléphoniques, ils ont pu citer en moyenne 2,21 services différents (é.t. = 1,9 ; min. = 0 ; max. = 10) qu'ils déclarent avoir déjà utilisés.

B Matériel

Le système utilisé pour cette expérience était une simulation du service *'Santiago'* (*Cf.* Figure 5-3). Son fonctionnement était identique à celui de la simulation utilisée pour l'expérience 1. Seule la stratégie de présentation des informations (facteur 1, voir ci-dessous) a été modifiée.

C Procédure

La procédure de l'expérience 4 était identique à celle de l'expérience 1. Les dialogues étaient identiques bien que la stratégie de présentation ait été adaptée (voir plus loin). Les consignes étaient les mêmes. Dans les deux cas, les évaluations de charge cognitive subjective étaient faites pour moitié avec *NASA-TLX* et pour moitié avec *'Workload Profile'*.

Une légère modification a cependant été apportée. Dans l'expérience 1, l'objectif était d'identifier l'effet de l'erreur dès le premier dialogue du participant avec le système. Les conditions expérimentales étaient donc appliquées dès le premier appel du participant. Pour l'expérience 4, l'intérêt se porte sur la *relation type-mode* qui peut opérer dans tous les cas, indépendamment des effets de la première prise de contact. Pour cette raison, dans cette expérience, les participants ont d'abord pu réaliser un essai de familiarisation (un dialogue complet) avant de rencontrer la première condition expérimentale à leur deuxième essai. Cela a permis de réduire l'impact d'un éventuel effet d'apprentissage.

Par ailleurs, pour le test de rappel en fin de séance, le participant ne répondait pas à l'écrit. En fin d'entretien, environ cinq minutes après le dernier appel, l'expérimentateur demandait au participant de lui donner le *jour*, la *date* et l'*heure* des deux rendez-vous pris au cours des appels expérimentaux (soit 6 items au total). La plupart des participants ont d'abord exprimé leur surprise (gémissements et expressions dédiées). Ils répondaient ensuite à voix haute.

D Protocole expérimental

Deux variables indépendantes étaient manipulées dans cette expérience.

- ***Facteur 1* : Stratégie de présentation des informations**

Quatre stratégies de présentation des informations ont été évaluées, conçues sur la base de l'analyse proposée dans la présentation des objectifs. Le tableau de la Figure 7-6 représente ces quatre stratégies.

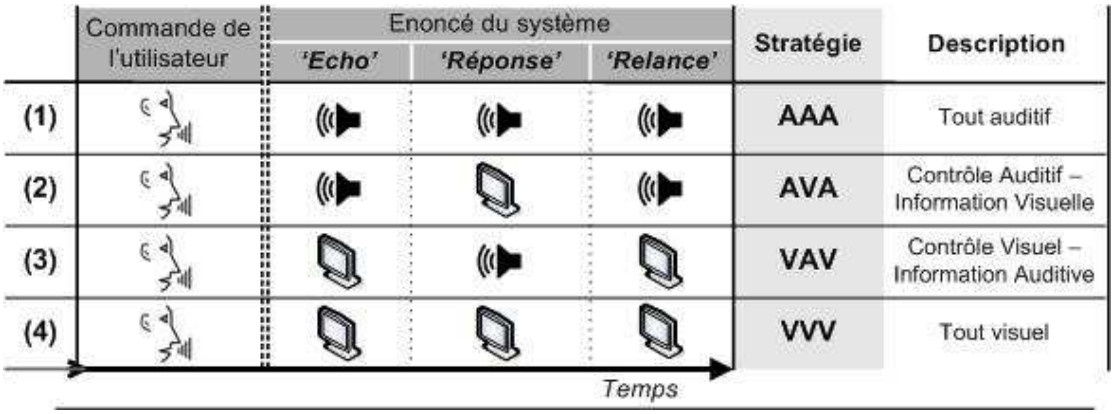


Figure 7-6 : Stratégie de présentation des informations (A: Auditif; V: Visuel)

La stratégie *'tout auditif'* (AAA) correspond au mode vocal, en vigueur dans le domaine des SVI, *i.e.* la communication téléphonique. Elle fournit le modèle de référence pour la conception des autres stratégies. Dans toutes les stratégies, l'*écho'* était présenté en premier, puis la *'réponse'*, puis la *'relance'*. Quand l'un de ces types d'information était présenté visuellement, un laps de temps correspondant au temps de lecture (mesuré en pré-test) était laissé avant de présenter les informations du type suivant. Les temps mesurés ont correspondu à une lecture deux fois plus rapide que celle de la synthèse vocale. De ce fait, la présentation visuelle des énoncés était plus rapide que la présentation auditive.

Il s'agit d'une *variable intergroupe*. Quatre groupes ont été constitués, correspondant aux quatre conditions. Ainsi, chaque participant a rencontré une seule stratégie de présentation au cours de l'expérience.

- ***Facteur 2* : Essai avec ou sans erreurs de reconnaissance vocale**

Chaque participant a réalisé trois dialogues. Le premier était un essai de familiarisation qui permettait d'intégrer le fonctionnement du système. Il n'est pas inclut aux résultats :

- ***'Essai sans erreur'*** : L'un des essais s'est déroulé sans perturbation. Toutes les commandes/demandes du participant étaient correctement prises en compte ;
- ***'Essai avec erreur'*** : Pour l'autre essai, ***une erreur*** *de reconnaissance vocale* était artificiellement introduite après la requête du participant.

Il s'agit d'une *variable intragroupe*. Chaque participant rencontrait les deux conditions au cours de deux essais successifs. La position du dialogue avec erreur était contre-balancée

entre les deux essais expérimentaux. (Le dialogue de familiarisation n'était pas inclut dans ce contre-balancement et se trouvait toujours en position 1).

E Mesures dépendantes

Les mesures utilisées dans la présentation des résultats sont indiquées dans le Tableau 7-2. Ces indicateurs sont présentés plus en détail à la fin du chapitre 6.

Tableau 7-2 : Indicateurs utilisés pour la présentation des résultats de l'expérience 4

Rappel	Nombre d'items rappelés parmi 9 (*'jour de la semaine'*, *'date'* et *'heure'* des trois rendez-vous).
Charge cognitive subjective	Evaluée avec le questionnaire *'NASA-TLX'* pour une partie des sujets et *'Workload Profile'* pour l'autre.
Nombre de tours de parole (TDP)	Nombre total de prises de parole par le participant au cours du dialogue.
Nombre de mots	Nombre total de mots prononcés par le participant au cours du dialogue.
Durée	Durée totale du dialogue (en secondes).
Mode de correction de l'erreur	(1) *'Correction'* ; (2) *'Annulation'* ou (3) *'Acceptation'* Le cas *'acceptation'* est nouveau par rapport à l'expérience 1. Dans l'expérience 4, deux participants n'ont pas jugé utile de corriger l'erreur de reconnaissance vocale. Ils l'ont acceptée.
Moment de la prise de parole	Catégorisée relativement à l'énoncé du système : (1) Dès l'écho, (2) Pendant la liste, (3) Après la relance.
Nb de mots par TDP	Nombre de mots moyen prononcés par TDP dans les différentes *phases* du dialogue.

7.2.3 Résultats

Le corpus final était composé de 160 dialogues au cours desquels 921 commandes des participants ont été prises en comptes pour un total de 6370 mots. Le nombre moyen de tours de parole (soit le nombre d'énoncés prononcés pour commander le système) était de XX pour le dialogue sans erreur de reconnaissance vocale et de XX pour le dialogue avec une erreur. Les analyses statistiques réalisées sur les données sont précisées relativement à chaque indicateur présenté.

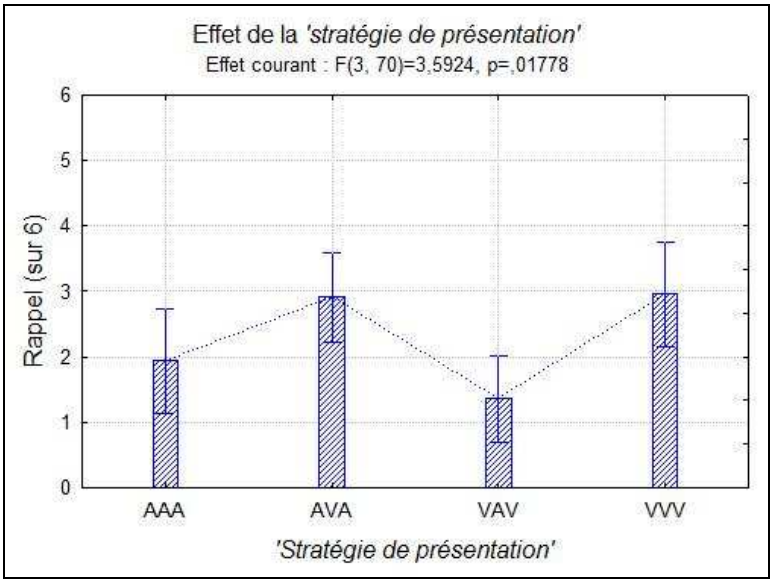
Deux corrélations intéressantes ont été observées. Le *nombre de tours de parole* dans le *dialogue sans erreur* de reconnaissance vocale était corrélé négativement au nombre d'heures passées sur internet par mois (***r* = -.28; *p* < .01**). La *durée du dialogue avec erreur* de reconnaissance vocale était corrélée négativement au nombre de logiciels cités par les participants comme fréquemment utilisés (***r* = -.22; *p* < .05**). Ces corrélations pourraient indiquer que la pratique en informatique des participants a eu une certaine influence sur leurs performances dans la tâche de DHM de l'expérience. Mais bien sur, aucune causalité ne peut être révélée par ce résultat.

A Indicateurs cognitifs généraux

Les indicateurs cognitifs généraux sont le rappel de la *date*, du *jour* et de l'*heure* des deux rendez-vous et les évaluations subjectives de charge cognitive (WP) ou de charge de travail (TLX).

- ***Rappel des rendez-vous pris au cours de l'expérience***

Les scores de rappel ont été analysés avec une procédure GLM incluant le *rappel* comme variable dépendante, la *stratégie de présentation* comme prédicteur catégoriel et le *nombre de tours de parole*, le *nombre de mots* produits et la *durée des dialogues* comme covariants.



Le test de Levene a indiqué que la distribution des scores de rappel était homogène *F(7,72)*=1,78 ; *p*=.10

La stratégie de présentation a eu un effet sur les scores de rappel :
***F(3,66)*=3,50 ; *p*<.05 ; η^2_p=.13**

Figure 7-7 : Performance de rappel des informations

On lit sur le graphique que la stratégie VAV a conduit à des scores de rappel plus faibles. Les comparaisons planifiées montrent que les scores obtenus dans cette stratégie (VAV) sont plus faibles que ceux obtenus dans la stratégie AVA (***F(1,66)*=8,90 ; *p*<.01 ; η^2_p=.11**) et dans la stratégie VVV (***F(1,66)*=7,49 ; *p*<.01 ; η^2_p=.10**). Les scores obtenus dans la stratégie entièrement auditive (AAA) n'étaient pas différents de la stratégie AVA (*F(1,66)*=2,75 ; *p*=.10 ; η^2_p=.04) ni de la stratégie VVV (*F(1,66)*=1,87 ; *p*=.17 ; η^2_p=.02).

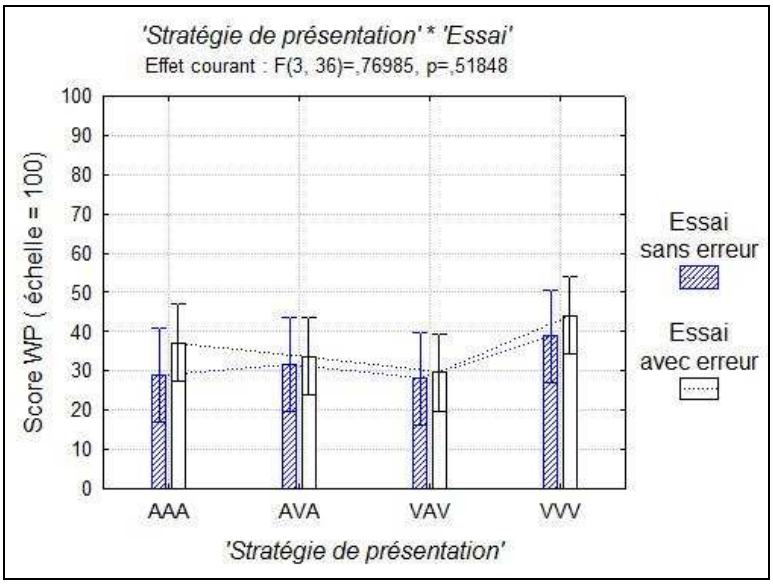
- ***Charge mentale / cognitive***

La procédure GLM utilisée pour les indices de charge cognitive incluait la stratégie de présentation comme variable indépendante et l'essai comme mesure répétée. Aucun covariant n'a été utilisé.

Les scores TLX n'ont pas révélé d'effet des facteurs expérimentaux (stratégie de présentation : *F(3,36)*=1,10 ; *p*=.36 ; η^2_p=.08 ; essai : *F(1,36)*=0,16 ; *p*=.68 ; η^2_p=.00 ; interaction : *F(3,36)*=0,42 ; *p*=.73 ; η^2_p=.03).

En revanche, les scores obtenus avec WP ont donné lieu à des différences. Pour ces scores, le test de Levene était marginalement significatif pour les scores obtenus à l'*essai sans erreur* (***F(3,36)*=2,99 ; *p*<.05**). A l'essai *'avec erreur'* la distribution des scores était homogène

(*F(3,36)*=0,70 ; *p*=.55). Les comparaisons de moyenne ont été appliquées malgré cette restriction.



La *stratégie de présentation* n'a pas eu d'effet sur les évaluations de charge :
F(3,36)=1,10 ; *p*=.35 ; η^2_p=.08

L'*essai* a eu un effet sur les évaluations de charge :
***F(1,36)*=5,80 ; *p*<.05 ; η^2_p=.13**

Et il n'y avait pas d'effet d'interaction :
F(3,36)=0,76 ; *p*=.51 ; η^2_p=.06

Figure 7-8 : Charge subjective aux deux essais

Les comparaisons planifiées ont montré que l'effet de l'essai était sensible uniquement pour la stratégie entièrement auditive (AAA) (***F(1,36)*=5,38 ; *p*<.05 ; η^2_p=.13**). La même comparaison pour les autres stratégies n'était pas significative. Ainsi, l'erreur a été jugée plus coûteuse quand le service était dans sa version « téléphonique classique ».

- ***Analyse discriminante à partir des scores WP***

L'analyse discriminante permet de prendre en compte simultanément l'ensemble des dimensions qui composent, par exemple, un questionnaire pour les projeter dans un espace composé d'autant de dimensions et tenter d'identifier un plan de coupe de cet espace qui permette de catégoriser correctement les groupes d'individus. Il s'agit de rechercher le meilleur profil du nuage de point formé par les données, *i.e.* celui qui discrimine le mieux les différents groupes.

Les résultats obtenus avec *'NASA-TLX'* n'ont pas permis de discriminer les participants des quatre groupes ayant utilisés les différentes *stratégies de présentation* (*Lambda Wilk* = 0,533; *F (18,88)* = 1,21; *p* = .26).

En revanche, les évaluations obtenues avec *'Workload Profile'* ont donné lieu à des résultats significatifs (*Lambda Wilk* = 0,207; *F (30,79)* = 1,88; *p* < .02). Les deux axes formant le plan de coupe étaient des compositions de plusieurs dimensions. L'axe principal était opposait le *traitement auditif* (.18) à la *réponse manuelle* (-.48). L'axe secondaire opposait le *sentiment de frustration* (.17) au *traitement perceptif/central* (-.46).

La Figure 7-9 illustre ces résultats. Sur l'axe principal (axe 1), la stratégie entièrement visuelle (VVV) était opposée aux trois autres. Dans cette stratégie, les participants ont indiqué une réponse manuelle plus importante. Sur l'axe secondaire (axe 2), la stratégie entièrement auditive (AAA) tendait à se détacher et à indiquer qu'un *traitement central* plus important était

nécessaire. Cette dimension du questionnaire peut être considérée comme l'indicateur de l'attention générale portée à la tâche, soit la concentration. Selon ce résultat, la stratégie auditive aurait imposé une plus grande concentration de la part des participants.

	Axe 1	Axe 2
T Perc./Centr.	-0,16	***-0,46***
T Réponse	-0,0	-0,18
T Spatial	-0,26	-0,28
T Verbal	-0,16	-0,15
T Visuel	-0,30	0,08
T Auditif	***0,18***	-0,31
T Manuel	***-0,48***	0,07
T Vocal	0,00	0,00
T Frustration	-0,11	***0,17***
T Perte de ctrl	-0,16	-0,10

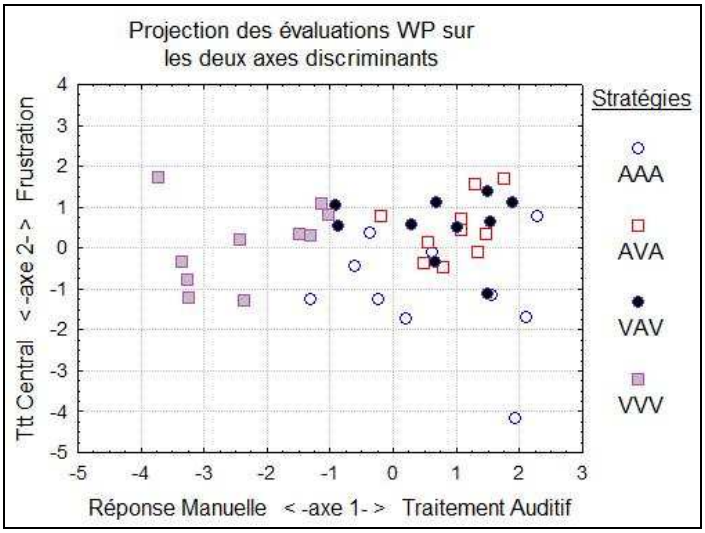


Figure 7-9 : Analyse discriminante avec WP (structure factorielle et graphique)

Les deux stratégies bimodales (AVA et VAV) ont donné lieu à des positionnements plus centraux. Ces participants étaient plus focalisés sur les processus perceptifs et semblent ont eu des profils de charge cognitive plus équilibrés.

- ***Effet du* 'mode de correction de l'erreur' *sur les scores WP***

Les indices généraux obtenus pour l'*essai avec erreur* ont été analysés en fonction du *'mode de correction'* de l'erreur utilisé par les participants (la distribution des modes de correction est présentée dans le Tableau 7-4 (p. 196), parmi les *indicateurs de comportement stratégique*). Pour cette comparaison, le *nombre de tours de parole*, le *nombre de mots* et la *durée du dialogue* ont été inclus en covariants.

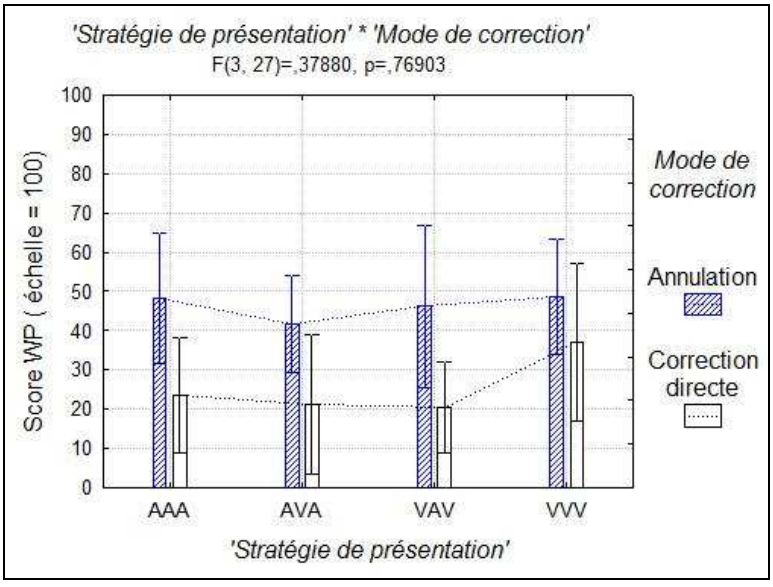


Figure 7-10 : Charge subjective pour l'essai avec erreur

Le test de Levene a indiqué une distribution homogène
F(7,30)=2,15 ; *p*=.06

La *stratégie de présentation* n'a pas eu d'effet sur les évaluations de charge :
F(3,27)=0,88 ; *p*=.46 ; η^2_p=.09

Le mode de correction a eu un effet :
***F(1,27)*=9,48 ; *p*<.01 ; η^2_p=.26**

Et il n'y a eu aucun effet d'interaction :
F(3,27)=0,37 ; *p*=.76 ; η^2_p=.04

Dans le dialogue avec erreur, la charge cognitive était évaluée comme plus importante par les participants qui ont corrigé par une *'annulation'* que par ceux qui ont fait une *'correction directe'*. Pour l'annulation, la moyenne brute était de 41,7 (é.t. = 10,6) comparé à 28,2 (é.t. = 14,8) pour la correction directe. On voit que l'indice de charge cognitive subjective montre une forte sensibilité à un indicateur de comportement stratégique. La *'correction directe'* coûtait un tour de parole de moins à l'utilisateur et permettait de raccourcir le dialogue de 14 secondes environ, en moyenne.

B Indicateurs de réalisation de la tâche

Comme dans l'expérience 1, les comparaisons réalisées sur le *nombre de tours de parole*, le *nombre de mots* et la *durée des dialogues* ont été analysés avec la même procédure GLM que le rappel. Pour chacun de ces indicateurs, les deux autres étaient inclus en covariants.

Les tests de Levene ont révélés un manque d'homogénéité des données pour le *nombre de tours de parole* et pour le *nombre de mots*. Les comparaisons sont présentées malgré ces restrictions.

Tableau 7-3 : Durée, nombre de mots et de tours de parole pour les deux dialogues

Essai	Indicateur	**AAA**	**AVA**	**VAV**	**VVV**	Tests de Levene
Sans erreur	***TDP***	4,9 (1,5)	3,95 (0,5)	5,3 (2,1)	4,55 (2,68)	***F(3,76)*=6,42; *p*<.01**
	Mots	26,4 (8,1)	27,0 (7,7)	33,5 (13,8)	40,3 (27,6)	***F(3,76)*=3,55; *p*<.05**
	Durée *(sec.)*	116,0 (15,3)	75,4 (24,6)	103,5 (22,0)	64,7 (22,9)	*F(3,76)*=0,56; *p*=.64
Avec erreur	***TDP***	7,35 (2,6)	6,45 (1,23)	7,35 (3,1)	6,2 (1,9)	***F(3,76)*= 5,99; *p*<.01**
	Mots	39,45 (16,6)	40,65 (12,6)	48,6 (17,6)	62,45 (38,2)	***F(3,76)*=13,62; *p*<.00**
	Durée *(sec.)*	142,4 (30,8)	112,5 (20,4)	118,6 (18,8)	99,0 (24,3)	*F(3,76)*=1,12; *p*=.34

Les comparaisons ont révélé que la *stratégie de présentation des informations* n'a pas eu d'effet sur le nombre de *tours de parole* (*F(3,72)*=2,11 ; *p*=.10 ; η^2_p=.08). Ce facteur a eu un effet sur le *nombre de mots* (***F(3,72)*=4,93 ; *p*<.01 ; η^2_p=.17**) et un effet très marqué sur la *durée des dialogues* (***F(3,72)*=24,38 ; *p*<.00 ; η^2_p=.50**).

L'*essai* a eu un effet que sur le *nombre de tours de parole* (***F(1,72)*=4,33 ; *p*<.05 ; η^2_p=.05**), mais pas sur le *nombre de mots* (*F(1,72)*=0,12 ; *p*=.72 ; η^2_p=.00) ; et il a également eu un effet sur la *durée des dialogues* (***F(1,72)*=4,17 ; *p*<.05 ; η^2_p=.05**).

Et les effets d'interaction n'étaient pas significatifs : nombre de *tours de parole* (*F(3,72)*=0,83 ; *p*=.47 ; η^2_p=.03) ; *nombre de mots* (*F(3,72)*=1,75 ; *p*=.16 ; η^2_p=.06) et *durée des dialogues* (*F(3,72)*=2,29 ; *p*=.08 ; η^2_p=.08).

On constate dans le Tableau 7-3 que deux *tours de parole* supplémentaires environ étaient nécessaires dans le dialogue avec erreur. Pour le *nombre de mots*, la stratégie entièrement visuelle (VVV) a rendu les participants plus verbeux dans les deux dialogues.

Pour la *durée des dialogues*, c'est surtout la présentation visuelle des *'réponses'* (stratégies AVA et VVV) qui a raccourcit les dialogues. On constate que la présentation entièrement visuelle donnait lieu aux dialogues les plus courts alors que la présentation entièrement auditive donnait lieu aux dialogues les plus longs. Ces résultats indiquent que, quel que soit le *type d'information* présenté, sa présentation auditive est un facteur de ralentissement du dialogue.

C Indicateurs comportementaux

Les résultats proposés concernant le comportement stratégique des participants sont le mode de correction de l'erreur, le moment de la prise de parole, ainsi que le nombre de tours de parole et le nombre de mots produits dans chaque phase du dialogue.

- ***Mode de correction de l'erreur***

Le mode de correction utilisé par les participants a également fait l'objet de distributions différentes (Tableau 7-4). Le cas de l'*acceptation* n'était pas apparu dans l'expérience 1. Il s'agit de participants qui n'ont pas corrigé l'erreur de reconnaissance vocale.

Tableau 7-4 : Mode de correction dans le *dialogue avec erreur* de reconnaissance

	AAA	**AVA**	**VAV**	**VVV**	X^2
Annulation	7	*15*	*4*	8	**X^2 (6) = 19,6; p < .001**
Correction directe	13	*5*	*16*	10	
Acceptation	0	0	0	*2*	

Les participants ont eu tendance à choisir l'*annulation* plus souvent quand les énoncés du système étaient présentés avec la stratégie AVA. Ils ont choisi plus souvent la *correction directe* avec les stratégies AAA et VAV.

Les deux cas d'*acceptation* de l'erreur de reconnaissance vocale sont apparus pour des participants confrontés à la stratégie entièrement visuelle (VVV). Il semble que ces participants ont été moins attentifs à la tâche qui leur était prescrite.

- ***Moment de la prise de parole***

Le Tableau 7-5 présente le moment de la prise de parole dans la *phase de choix* (*Cf.* Figure 5-3). Les résultats présentés portent sur le *dialogue avec erreur*, lors du premier passage dans cette phase – au cours duquel l'erreur doit être corrigée – et lors du dernier passage – pour le choix final d'un rendez-vous. Dans le tableau, la notion *« d'étape »* est utilisée pour faire la distinction entre ces deux passages dans la même phase. Les tests X^2 pratiqués sur ces deux distributions sont très significatifs et indiquent que la *stratégie de présentation* a eu une influence sur le moment de la prise de parole dans cette phase du dialogue.

Tableau 7-5 : Moment de la prise de parole dans le *dialogue avec erreur*

Etape	Prise de parole	AAA	AVA	VAV	VVV	X^2
Correction de l'erreur	*Dès l'écho*	*14*	*2*	5	5	**X^2 (6) = 53,7; p < .000**
	Pendant la liste	5	*0*	*13*	3	
	Après la relance	*1*	*18*	*2*	11	
Choix	*Pendant la liste*	13	0	14	6	**X^2 (3) = 26,5; p < .000**
	Après la relance	7	20	6	14	

On lit dans le Tableau 7-5 que les stratégies dans lesquelles les *'réponses'* étaient présentées visuellement (AVA et VVV) ont incité les participants à attendre la fin de la liste avant de prendre la parole. Quand, au contraire, les *'réponses'* étaient présentées sur la modalité auditive (AAA et VAV), les prises de parole étaient plus précoces et apparaissaient fréquemment pendant la présentation de la liste de *'réponses'* par le système. De plus, pour la correction de l'erreur de reconnaissance vocale, la stratégie entièrement auditive (AAA) a incité les participants à faire la correction immédiatement après l'*écho'*. Ainsi, la présentation auditive des informations a forcé les participants à une réaction plus immédiate.

- ***Nombre de mots à différentes étapes du dialogue***

La Figure 7-11 présente le nombre de mots prononcés par tour de parole par les participants. Les valeurs présentées sont des moyennes car chaque participant peut avoir franchi l'étape indiquée en un ou plusieurs *tours de parole*.

Les tests de Levene ont révélé que les distributions étaient homogènes pour l'étape de requête (*F(3,75)*=0,84 ; *p*=.47) et pour l'étape de choix (*F(3,75)*=1,76 ; *p*=.16). En revanche, le test était significatif pour l'étape de choix (***F(3,27)*=14,10 ; *p*<.00**). Les résultats de l'analyse sont cependant présentés.

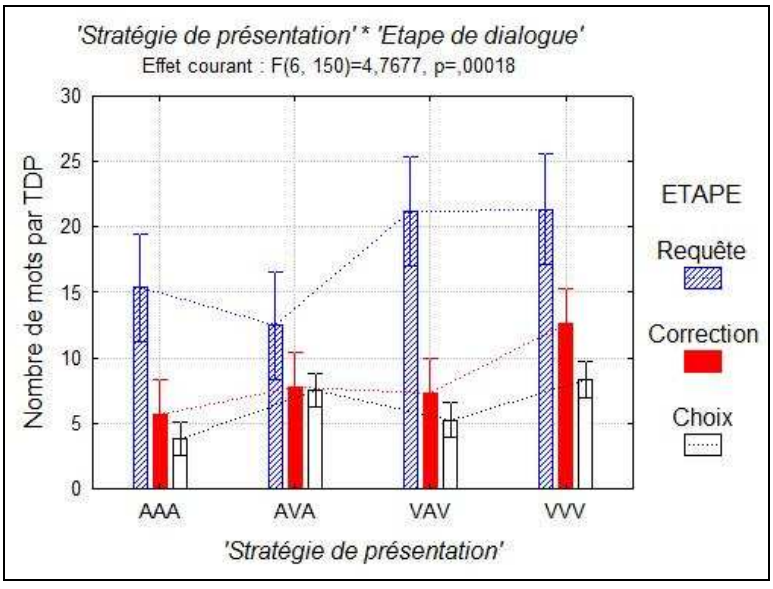


Figure 7-11 : Nombre de mots par TDP à différentes étapes

La *stratégie de présentation* a eu un effet significatif sur le nombre de mots par tour :

***F(3,75)*=4,99 ; *p*<.01 ; η^2_p=.16**

Le nombre de mots par TDP était très différent entre les *étapes du dialogue* :

***F(2,150)*=108,3; *p*<.00; η^2_p=.59**

Et il y avait un effet d'interaction entre ces deux facteurs :

***F(6,150)*=4,76 ; *p*<.00 ; η^2_p=.16**

On lit sur le graphique que le nombre de mots prononcés par tour était plus important lors de la requête et plus faible lors de la correction de l'erreur et du choix final. L'effet de la *stratégie*

de présentation n'était pas le même aux différentes étapes. Pour la *requête*, les *'relances'* visuelles (stratégies VAV et VVV) avaient tendance à rendre les participants plus verbeux. Pour les étapes de *correction* et de *choix*, c'est le mode de présentation des *'réponses'* qui permet de distinguer les *stratégies*. Les *'réponses'* auditives ont amené les participants à produire des commandes plus courtes ; et c'est surtout face à la stratégie entièrement visuelle (VVV) que les participants sont restés verbeux dans ces étapes avancées du dialogue.

On peut noter que la stratégie AVA est celle qui a conduit au meilleur équilibre entre les différentes étapes.

7.2.4 Conclusion

On a constaté dans cette expérience que l'une des stratégies bimodales, la stratégie VAV (*contrôle visuel – information auditive*), conduisait à une dégradation de la performance de rappel, alors que cette baisse de performance n'était pas sensible avec la stratégie entièrement auditive (AAA) correspondant au mode téléphonique classique.

Ce résultat est contraire à l'effet de suffixe, propre à la stratégie AAA, constaté dans l'expérience 3. La différence entre ces résultats peut être expliquée par la différence entre les consignes des expériences. La consigne de rappel explicite de l'expérience 3 a incité les participants à mettre en place une stratégie d'autorépétition des informations. Les *'relances'* auditives les ont gênés dans l'exécution de cette stratégie, ce qui a conduit à l'effet de suffixe. Dans l'expérience 4, aucune consigne de rappel n'a conduit à la mise en place de cette stratégie et la dégradation de la performance dans cette stratégie entièrement auditive n'a pas eu lieu. Le rappel faible obtenu dans la stratégie VAV doit être expliqué autrement.

Le point de vue proposé dans le cadre de la *théorie de la charge cognitive* est basé sur la quantité d'information à traiter et le dépassement de la capacité de la mémoire de travail. Selon ce point de vue, l'utilisation des canaux visuel et auditif pour présenter des informations distinctes permet au sujet apprenant d'accroître sa capacité disponible en mémoire de travail puisque deux processeurs sont disponibles pour traiter l'information (voir par exemple, Tindall-Ford, Chandler & Sweller, 1997, p. 283). Ces auteurs notent que des exceptions pourraient cependant être trouvées si le matériel auditif est structuré de telle sorte qu'il surcharge la mémoire de travail ; si, par exemple, la longueur des éléments présentés auditivement est trop importante, ou s'il existe une complexité inhérente au contenu présenté (comme c'est le cas lors de la description d'une figure géométrique), ou encore si le matériel présenté auditivement est redondant avec celui qui est présenté visuellement (Tindall-Ford, Chandler & Sweller, 1997, p. 284). Or, le matériel utilisé dans l'expérience 4 ne correspond à aucun de ces cas. La longueur du matériel présenté auditivement était plus faible dans la stratégie VAV que dans la stratégie AAA. Dans toutes les conditions, la complexité était faible dans la mesure où il s'agissait simplement de prendre un rendez-vous selon le système du calendrier occidental, très largement connu de tous les participants. Et aucune redondance

n'était utilisée pour présenter les informations. Par ailleurs, la contiguïté temporelle entre les éléments demeurait la même d'une stratégie à l'autre, ce qui pouvait laisser supposer qu'aucun effet de partage de l'attention ne devait apparaître dans une stratégie plus que dans les autres. En outre, les évaluations subjectives de charge cognitive n'ont pas révélé d'augmentation de la charge dans la stratégie VAV. La moyenne était au contraire la plus faible pour les participants qui ont été confrontés à cette stratégie, quoiqu'il n'y ait pas eu de différence significative. Pour ces différentes raisons, on ne peut pas supposer que la dégradation de la performance de rappel ait été due à une surcharge de la mémoire de travail dans cette expérience. Une autre explication doit être recherchée.

L'*effet de préemption* (Helleberg & Wickens, 2003) correspond au captage de l'attention que l'on constate lors de la présentation auditive de l'information, qui tend à interrompre l'activité en cours. Dans la stratégie AVA, qui a donné lieu à un bon rappel, les *informations cibles* étaient présentées visuellement et pouvaient être lues par les participants. Les informations relatives à la tâche de *contrôle du dialogue*, présentées sur la modalité auditive venaient alors préempter l'activité d'exploration visuelle pour rappeler au participant quelles possibilités d'action il avait. Dans ce cas, l'effet de préemption n'a pas été négatif puisque le rappel était bon dans cette condition. La stratégie VAV correspondait au cas opposé. Les informations de contrôle du dialogue étaient présentées visuellement. Le participant voyait d'abord l'*écho'* de sa commande s'afficher, ce qui orientait son attention visuelle sur la tâche de contrôle du dialogue, puis la présentation auditive des *'réponses'* préemptait son attention pour l'orienter sur la tâche cible, puis il devait à nouveau réorienter son attention visuelle, sur les *'relances'*. Dans ce cas, l'effet de préemption était négatif car il était mal utilisé. L'*écho'* visuel était rapidement consulté. L'effet de préemption captait alors toute l'attention du participant qui se mettait souvent à réagir à chaque proposition (le Tableau 7-5 indique que dans la stratégie VAV les participants tendaient à prendre la parole plus fréquemment *« pendant la liste »*, *i.e.* pendant la présentation des *'réponses'*). La présentation visuelle de la *'relance'*, après la fin de la liste, arrivait alors après cette réorientation de l'activité et obligeait les participants à une seconde réorientation, vers la *tâche de contrôle*. Ainsi, lors de la présentation de chaque énoncé complet du système (*'écho'*, *'réponse'* et *'relance'*) le participant devait alterner entre les deux types d'activités nécessaires à la bonne conduite du dialogue, ce qui a provoqué un effet de partage de l'attention. On voit que c'est l'alternance entre les deux types d'activité qui a induit une concurrence entre les deux types de traitement, ce qui s'est concrétisé en un partage de l'attention. Cet effet ne s'explique pas, alors, par une surcharge de la mémoire, selon une *métaphore énergétique*, mais par un guidage inadapté des actes des participants qui ne permettait pas la *« résonance »* des deux types d'activité, selon une *métaphore vibratoire*. Les deux types d'activité impliqués correspondent aux deux fonctions dialogiques principales, selon la catégorisation de Bunt, qui ont été utilisées dans la conception du matériel expérimental pour mettre en relation les *types d'information* à présenter et les *modalités de présentation*. Dans ce sens, il semble que l'on puisse conclure qu'il existe une relation *type-mode* spécifique et qu'elle trouve l'une de ses causes dans l'*effet de préemption*.

Comme dans les expériences précédentes, les résultats obtenus étaient liées aux variations comportementales induites par les stratégies de présentation. Notamment, l'analyse discriminante faite sur les évaluations des dimensions du questionnaire *'Workload Profile'* a permis de comprendre que la *'réponse manuelle'* face au système a été plus coûteuse pour les participants qui étaient confrontés à la stratégie entièrement visuelle (VVV).

Du point de vue de la performance, les gains de temps sont évidents quand les *'réponses'* sont présentées visuellement (stratégies AVA et VVV), mais la trop grande verbosité des participants dans la stratégie entièrement visuelle (VVV) la disqualifie. Ces participants ont eu tendance à formuler de longues demandes dans toutes les phases du dialogue. Pour un système de dialogue, cela peut provoquer un plus grand nombre d'erreurs d'interprétation. Par ailleurs, cela indique que ces participants étaient moins focalisés sur la tâche. Quoiqu'il en soit, les participants dans cette stratégie ont eu une bonne performance de rappel, ce qui permet de supposer que la présentation visuelle des *'réponses'* a de l'importance pour la mémorisation.

7.3 Expérience 5 : Quel niveau d'analyse de la spécificité modale ?

7.3.1 Objectifs et hypothèses

Les résultats obtenus dans les expériences 3 et 4 ont permis de mettre en évidence la spécificité de la relation *type-mode*. Celle-ci repose sur une catégorisation fonctionnelle des contenus présentés à l'utilisateur. On peut se demander si une catégorisation plus fine pourrait permettre une amélioration plus conséquente du processus dialogique. D'une part, les trois types d'information proposés (*'écho'*, *'réponse'* et *'relance'*) sont composés de plusieurs éléments qui peuvent être décomposés. D'autre part, chaque élément identifié peut avoir plusieurs rôles eu égard aux différentes fonctions dialogiques. Pour ces raisons, il est possible de proposer des stratégies de présentation qui pourraient permettre une plus grande efficacité dans le dialogue, mais dont la conception est plus complexe car elle repose sur une analyse plus riche.

En conséquence, l'objectif de l'expérience 5 était de comparer une stratégie de présentation plus riche, conçue selon cette analyse, avec la stratégie AVA dont on a vu qu'elle permettait une bonne performance et avec la stratégie entièrement auditive (AAA) qui servait de condition contrôle. Il était nécessaire d'utiliser un service présentant une plus grande quantité d'information de façon à permettre la conception d'une *« stratégie multimodale riche »*. Le service *'Cinéliste'* a été imaginé dans cette perspective, car il permet la présentation d'un grand nombre de champs d'information (*'réalisateur'*, *'acteurs'*, *'genre'*, *'année'*, etc.) qui peuvent avoir une nature différente (*e.g.* des personnes, des dates, l'histoire du film). Cette plus grande diversité permet d'aller plus loin dans l'analyse.

La *'stratégie multimodale riche'* conçue pour l'expérience était basée sur la décomposition des éléments et la présentation soit auditive, soit visuelle soit bimodale de chacun des éléments. Les fonctions dialogiques n'étaient pas attribuées exclusivement l'une de l'autre, mais indépendamment :

- L*'écho'* de la demande de l'utilisateur a une fonction d'alerte qui est liée au contrôle du dialogue (*fonction de feedback, Cf.* Figure 2-4, p. 57) et il a une *fonction d'information orientée interaction* (nécessité de confirmer ce qui a été compris). La coexistence de ces deux fonctions a justifié une présentation bimodale de ce type d'information.
- Les *'réponses'* ont été structurées sous la forme d'une liste visuelle pour les champs d'information orientés tâche (*'réalisateur'*, *'titre'*, *'genre'*, *'nationalité'*, *'année'*, *'acteur 1'*, *'acteur 2'*) et d'une diffusion auditive du *'synopsis'* du film (20 à 22 mots). En effet, le *'synopsis'* a un contenu narratif qui peut relever d'une *fonction de structuration du discours* et/ou d'une *fonction de feedback* (deux des *fonctions de contrôle*). Ainsi, les *'réponses'* étaient scindées en deux types auxquels les modes de présentation étaient attribués exclusivement.

- Dans le cas des *'relances'*, des adaptations du contenu ont été apportées spécifiquement à chaque mode. Les *'relances'* consistent d'une part à présenter la liste des commandes disponibles, ce qui correspond à une *fonction d'information orientée interaction* et justifie l'utilisation du mode visuel. D'autre part, il s'agit de prévenir l'utilisateur – sur un mode coopératif – que les commandes disponibles sont affichées (*fonction de gestion de l'interaction : gestion des obligations sociales*).et de lui indiquer qu'une réponse est attendue (*fonction de gestion de l'interaction : gestion des tours*), soit l'énoncé : *« Les commandes disponibles sont affichées à l'écran. Je vous écoute. »*

Ici, la notion de *« stratégie multimodale »* désigne cette combinaison riche par opposition à la *« stratégie bimodale »* (AVA) conçue sur un mode exclusif. L'hypothèse était que la stratégie multimodale pourrait permettre une interaction de meilleure qualité et une exploration des données plus facile qui devait se traduire en un gain de performance dans le dialogue et au test de rappel.

7.3.2 Méthode

A Participants

Le nombre final de participants était de 48 (4 hommes, 44 femmes). La moyenne d'âge du groupe était de 20,78 ans (*é.t.* = 3,03 ; *min.* = 19,25 ; *max.* = 40,25). Les participants étaient des étudiants en psychologie à l'Université Rennes 2 en Licence – deuxième année. Ils participaient à l'expérience dans le cadre des enseignements en psychologie expérimentale, en tant qu'activité obligatoire associée au cours.

Le niveau d'expérience en informatique des participants était moyen. Ils ont déclaré passer en moyenne 12,83 heures par mois à utiliser un ordinateur (*é.t.* = 13,1 ; *min.* = 1 ; *max.* = 70) dont 6,93 sur internet (*é.t.* = 7,65 ; *min.* = 0 ; *max.* = 30). Pour comparaison, ils ont déclaré passer 11,75 heures par mois à regarder la télévision (*é.t.* = 8,15 ; *min.* = 0 ; *max.* = 35) et 9,27 heures par mois à lire (revues et/ou littérature : *é.t.* = 8,20 ; *min.* = 0 ; *max.* = 42).

B Matériel

Le système utilisé pour cette expérience était une simulation du service *'Cinéliste'* (*Cf.* Figure 5-4). Son fonctionnement était identique à celui de la simulation utilisée pour l'expérience 2. Seule la stratégie de présentation des informations (facteur 1, voir ci-dessous) a été modifiée.

C Procédure

La procédure de l'expérience 5 était strictement identique à celle de l'expérience 2 qui a été présenté pages 163-164.

D Protocole expérimental

Deux variables indépendantes ont été manipulées dans cette expérience.

- ***Facteur 1*: Stratégie de présentation des informations**

Trois conditions ont été évaluées :

1. ***'Vocale'*** : Cette condition correspond à la condition de *'style normal'* de l'expérience 2 et s'apparente à la communication téléphonique classique ;
2. ***'Bimodale'*** : Cette condition correspond à la stratégie AVA de l'expérience 4 (condition 2), dont les résultats ont montré qu'elle était la plus favorable à une interaction performante ;
3. ***'Multimodale'*** : Cette condition correspond à la stratégie de présentation présentée dans les objectifs de l'expérience, ci-avant. Elle est basée sur une sous-catégorisation des types d'information. Les énoncés sont mieux synchronisés et intègrent certaines références croisées entre les modes de présentation.

Il s'agit d'une *variable intergroupe*. Trois groupes ont été constitués, correspondant aux trois conditions. Chaque participant a rencontré une seule de ces conditions au cours de l'expérience.

- ***Facteur 2* : Essai**

Chaque participant a réalisé trois dialogues. Le premier dialogue correspondait à un essai de familiarisation avec le système et n'est pas intégré aux résultats expérimentaux. Le second dialogue avait une consigne simple, le troisième avait une consigne complexe :

- ***'Essai simple'*** : Le participant avait pour consigne d'identifier un film et d'en demander la lecture. La question était : « Dans quel film Madame Baines meurt-elle ? » ;
- ***'Essai complexe'*** : Le participant avait pour consigne d'identifier deux films et de demander la lecture de l'un des deux. La question était : *« Identifiez le film le plus ancien et le plus récent parmi les films de Fritz Lang et demandez la lecture de l'un d'eux. »*

Il s'agit d'une *variable intragroupe*. Chaque participant rencontrait les deux conditions au cours de deux essais successifs.

E Mesures dépendantes

Les mesures utilisées dans la présentation des résultats sont indiquées dans le Tableau 7-6. Lorsque des explications sur la signification des différents indicateurs sont nécessaires, elles sont précisées lors de la présentation des résultats.

Tableau 7-6 : Indicateurs utilisés pour la présentation des résultats de l'expérience 5

Rappel	Nombre d'items (mots) correctement rappelés parmi 30 (informations sur le film : *'titre'*, *'année', 'genre', 'nationalité', 'acteur 1', 'acteur 2' ;* et histoire du film : *'synopsis'*).
Charge cognitive subjective	Evaluée avec le questionnaire *'Workload Profile'*.
Nombre de tours de parole (TDP)	Nombre total de prises de parole par le participant au cours du dialogue.
Nombre de mots	Nombre total de mots prononcés par le participant au cours du dialogue.
Durée	Durée totale du dialogue (en secondes).
Nombre d'écoutes des informations	Nombre de consultations du ou des film(s) identifié(s) par le participant au cours du dialogue.
Temps de réponse dans la *'phase de choix'*	Ce temps correspond au temps de consultation du film choisi par le participant avant d'en demander la lecture.
Mode d'exploration de la liste	(1) Direct, (2) Indirect. (Voir détail dans la présentation des résultats).
Moment de la prise de parole	Catégorisé relativement à l'énoncé du système : (1) *'Pendant les réponses'*, (2) *'Pendant le menu'*, (3) *'Après la relance'*.
Taux de recouvrement	Entre parole du participant et parole du système. Ce taux correspond au pourcentage de prises de parole du participant au cours desquelles un temps de recouvrement est apparu.

7.3.3 Résultats

Le corpus final est composé de 96 dialogues au cours desquels 1866 commandes des participants ont été prises en compte pour un total de 2642 mots. Le nombre moyen de tours de parole (soit le nombre d'énoncés prononcés pour commander le système) était de 14,8 pour le dialogue simple et de 24,1 pour le dialogue complexe. Les analyses statistiques réalisées sur les données sont précisées relativement à chaque indicateur présenté.

Les transcriptions des dialogues ont été faites sur le même modèle que celles des dialogues de l'expérience 2. Les analyses statistiques réalisées sur les données sont présentées relativement à chaque indicateur proposé.

A Indicateurs cognitifs généraux

La performance de rappel a été évaluée sur les mêmes bases que dans l'expérience 2. Les procédures GLM utilisées pour la comparaison des scores de rappel et pour la comparaison des indices de charge cognitive est également est également identique à celle de l'expérience 2.

Tableau 7-7 : Rappel et charge cognitive subjective pour les deux dialogues

Essai	Indicateur	**Vocal**	**Bimodal**	**Multimodal**	Test de Levene
Simple	***Rappel***	17,4 (5,6)	16,1 (6,9)	15,6 (7,6)	*F(2,45)*=0,90; *p*=.41
	Charge (WP)	33,6 (12,4)	37,2 (16,6)	31,6 (16,2)	*F(2,45)*=1,11; *p*=.33
Complexe	***Rappel***	11,4 (3,6)	12,6 (4,5)	11,7 (4, 3)	*F(2,45)*=1,03; *p*=.36
	Charge (WP)	36,1 (17,0)	36,1 (16,3)	35,6 (12,0)	*F(2,45)*=0,65; *p*=.52

La '*stratégie de présentation des informations'* n'a pas eu d'effet sur le score de rappel (*F(2,39)*=0,21 ; *p*=.80 ; η^2_p=.01), ni sur les évaluations de charge cognitive (*F(2,39)*=0,25 ; *p*=.77 ; η^2_p=.01).

Le facteur *'essai'* n'a pas eu d'effet sur le score de rappel (*F(1,39)*=2,53 ; *p*<.11 ; η^2_p=.06), ni sur les évaluations de charge cognitive (*F(1,36)*=0,18 ; *p*=.66 ; η^2_p=.00).

Il n'y avait pas d'effet d'interaction entre les facteurs pour le rappel (*F(2,39)*=0,22 ; *p*=.79 ; η^2_p=.01), mais il y avait un **effet d'interaction marginal** pour les évaluations de charge cognitive (*F(2,39)*=2,77 ; *p*=.07 ; η^2_p=.12). Mais aucune des comparaisons planifiées n'a atteint la significativité.

Ainsi, les facteurs expérimentaux n'ont pas eu d'effet sur les indicateurs cognitifs généraux.

B Indicateurs de réalisation de la tâche

Les procédures GLM utilisées pour comparer les indicateurs de réalisation de la tâche sont identiques à celles utilisées dans l'expérience 2. Le test de Levene a indiqué un manque d'homogénéité des données pour le nombre mots produits à l'essai complexe.

Tableau 7-8 : Durée, nombre de mots et de tours de parole pour les deux dialogues

Essai	Indicateur	**Vocal**	**Bimodal**	**Multimodal**	Test de Levene
Simple	***TDP***	15,4 (4,3)	13,6 (5,4)	15,25 (5,8)	*F(2,45)*=0,72; *p*=.48
	Mots	20,6 (7,4)	23,5 (12,2)	25,5 (11,0)	*F(2,45)*=1,13; *p*=.33
	Durée *(sec.)*	421,6 (114)	209,4 (60)	241,9 (69)	*F(2,45)*=3,05; *p*=.06
Complexe	***TDP***	24,1 (6,5)	22,8 (5,7)	25,1 (7,3)	*F(2,45)*=0,10; *p*=.90
	Mots	27,5 (6,9)	31,8 (11,2)	36,0 (18,2)	***F(2,45)*=7,42; *p*<.01**
	Durée *(sec.)*	500,3 (152)	303,8 (91)	304,0 (83)	*F(2,45)*=2,71; *p*=.07

La *stratégie de présentation des informations* n'a pas eu d'effet sur le *nombre de tours de parole* (*F(2,41)*=0,12 ; *p*=.88 ; η^2_p=.00) ni sur le *nombre de mots* (*F(2,41)*=1,19 ; *p*=.31 ; η^2_p=.05). En revanche, ce facteur a eu un effet très significatif sur la *durée des dialogues* (***F(2,41)*=30,21 ; *p*<.00 ; η^2_p=.59**).

L'*essai'* a eu un effet sur le *nombre de tours de parole* (***F(1,41)*=10,41 ; *p*<.01 ; η^2_p=.20**), plus important à l'essai complexe. Il n'a eu aucun effet sur le *nombre de mots* prononcés (*F(1,41)*=0,00 ; *p*=.97 ; η^2_p=.00). Et il a eu un effet sur la *durée des dialogues* (***F(2,41)*=9,14 ; *p*<.01 ; η^2_p=.18**).

Un **effet d'interaction marginal** est apparu entre les deux facteurs expérimentaux pour le *nombre de tours de parole* prononcés (*F(2,41)*=2,62 ; *p*=.08 ; η^2_p=.11). L'augmentation du nombre de tours de parole à l'essai complexe était plus important avec la *'stratégie de présentation vocale'*. Cet effet n'est pas apparu pour le nombre de mots (*F(2,41)*=0,55 ; *p*=.57 ; η^2_p=.02), ni pour la durée (*F(2,41)*=0,38 ; *p*=.68 ; η^2_p=.01).

Ainsi, les stratégies de présentation bimodale et multimodale ont seulement permis aux participants de gagner du temps dans les dialogues.

C Indicateurs comportementaux

Comme dans l'expérience 2, les résultats portant sur le comportement stratégique des participants sont le *'nombre de consultations'* des informations sur les films, le *'mode d'exploration'* de la liste des réponses, le *'moment de la prise de parole'* par les participants et le *'taux de parole avec recouvrement'*. De plus, un indicateur portant sur le type de vocabulaire utilisé a été ajouté : le *'taux d'utilisation des titres de films dans les commandes'*.

- ***Nombre de consultations des informations sur les films***

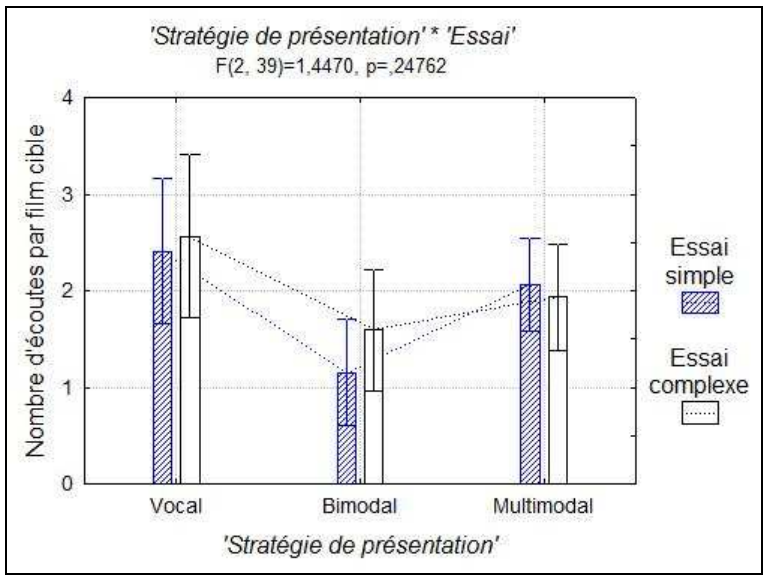


Figure 7-12 : Nombre d'écoutes des films cible

Les tests de Levene ont révélé une distribution non homogène :
<u>simple</u> : ***F(2,45)*=11,72 ; *p*<.01**
<u>complexe</u> : *F(2,45)*=1,04 ; *p*=.35

Le *'style d'énoncé'* a eu un effet sur le nombre d'écoutes :
***F(2,39)*=3,58 ; *p*<.05 ; η^2_p=.15**

L'essai n'a pas eu d'effet :
F(1,39)=1,09 ; *p*=.30 ; η^2_p=.02

Et on ne constate aucun effet d'interaction entre les deux facteurs :
F(2,39)=1,44 ; *p*=.24 ; η^2_p=.06

Pour le nombre d'écoute des films, la procédure GLM appliquée était également la même que dans l'expérience 2.

La *stratégie entièrement auditive* est celle qui a conduit au plus grand nombre de consultations des informations. La *stratégie bimodale*, dans laquelle les *'réponses'* étaient présentées uniquement sur la modalité visuelle a donné lieu au nombre de consultation le plus faible. La *stratégie multimodale*, qui présentait une partie des *'réponses'* visuellement et une partie auditivement a fait l'objet d'un nombre de consultation intermédiaire. Ainsi, plus la quantité d'informations présentées auditivement était importante, plus les participants tendaient à réécouter les informations.

- ***Temps de réponse dans la phase de choix***

Le temps de réponse dans la phase de choix correspond au temps de lecture pris par le participant après avoir identifié le (ou l'un des) film(s) recherché(s) et avant la commande suivante. Dans le cas de la *stratégie vocale*, ces temps étaient proches de zéro et ces résultats n'ont pas été relevés au cours de la transcription des résultats. Le Tableau 7-9 présente les résultats pour les stratégies *'bimodale'* et *'multimodale'*.

Tableau 7-9 : Temps de réponses dans la *'phase de choix' (en secondes)*

Essai	Film	**Bimodale**	**Multimodale**	Test de Levene
Simple	***Un seul film***	21,3 (14,6)	12,1 (12,8)	*F(1,30)*=1,11 ; *p*=.29
Complexe	***Film 1***	26,6 (17,2)	11,6 (11,5)	*F(1,30)*=2,36 ; *p*=.13
	Film 2	30,6 (22,1)	13,8 (13,6)	***F(1,30)*=5,49 ; *p*<.05**

Dans le dialogue simple, les temps de réponse ont été plus courts avec la *stratégie multimodale* en comparaison à la *stratégie bimodale*. La différence était significative (***F(1,29)*=4,26 ; *p*<.05 ; η^2_p=.12**).

Dans le dialogue complexe, les temps de réponse ont également été plus courts avec la *stratégie multimodale* en comparaison à la *stratégie bimodale*. Là aussi, la différence était significative (***F(1,29)*=10,13 ; *p*<.01 ; η^2_p=.25**).

- ***Mode d'exploration***

Dans l'expérience 2, deux modes d'exploration (direct *vs.* indirect) ont été proposés pour décrire les cheminements des participants dans les dialogues. Les stratégies de présentation bimodales ont conduit à une variabilité plus importante des parcours. Le mode « indirect » a été divisé en ses sous-catégories : « balayage », « picorage » et « contrôle » (*Cf.* présentation des modes d'exploration dans les résultats de *l'expérience 2*, p. 169).

Seul le *dialogue simple* a donné lieu à une différence significative entre les stratégies. Dans ce dialogue, on constate qu'avec la *stratégie vocale*, l'exploration *directe* était plus fréquente qu'avec les autres stratégies. La *stratégie bimodale* a incité les participants à explorer la liste par *picorage*, *i.e.* à choisir les films consultés de façon aléatoire. Dans la *stratégie*

multimodale, l'exploration par balayage permettait aux participants de revoir la liste des films après chaque consultation des détails pour avancer dans la liste.

Tableau 7-10 : Modes d'exploration de la liste de films

Essai	Mode d'exploration	**Vocal**	**Bimodal**	**Multimodal**	X^2
Simple	***direct***	7	5	3	***$X^2(2)$* = 11,14 ; *p<.05***
	balayage	2	1	8	
	picorage	7	10	5	
Complexe	***direct***	10	10	8	*$X^2(2)$* = 3,48 ; *N.S.*
	balayage	3	5	7	
	contrôle	3	1	1	

Dans le dialogue complexe, la consigne qui était donnée aux participants d'identifier deux films différents les a incité à privilégier le mode d'exploration *direct* car la comparaison entre les films était plus facile dans ce cas.

- ***Moment de la prise de parole***

Comme dans l'expérience 2, les moments des prises de parole présentés portent sur le *dialogue simple* dans la *phase de détail*. Le Tableau 7-11 présente des pourcentages calculés à partir des différents passages dans cette phase, dont le nombre était variable en fonction des initiatives des participants (moy. = 7,81 ; é.t. = 2,03 ; min. = 1 ; max. = 11).

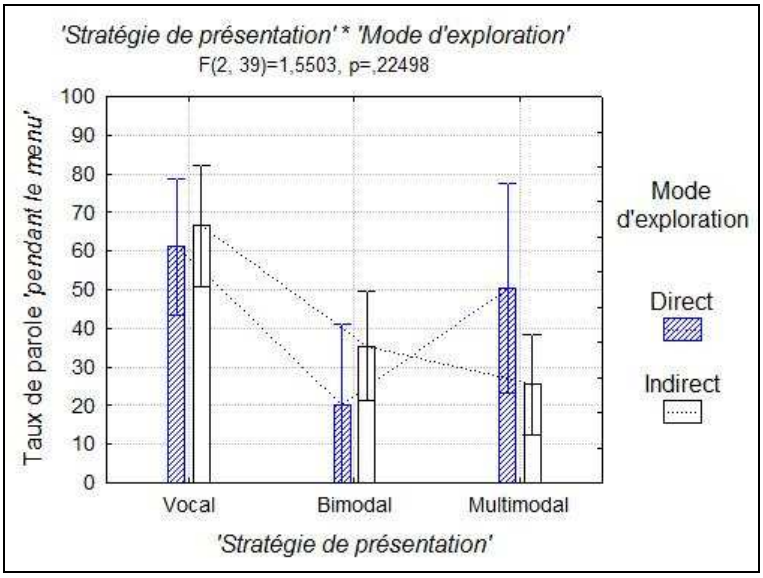
Tableau 7-11 : Distributions des moments de prises de parole dans la *'phase de détail'*

Moment	**Vocal**	**Bimodal**	**Multimodal**	Test de Levene
Pendant les *'réponses'*	3,2 (6,4)	23,9 (24,3)	11,2 (16,2)	***F(5,42)*=9,97 ; *p*<.00**
Pendant le menu (*'relance'*)	64,2 (40,8)	30,5 (27,2)	30,2 (24,9)	*F(5,39)*=2,06 ; *p*=.08
Après la *'relance'*	32,5 (42,4)	45,5 (36,2)	58,5 (32,6)	***F(5,42)*=2,46 ; *p*<.05**

Comme dans l'expérience 2, seuls les pourcentages de prises de parole correspondants aux réponses *'pendant le menu'* sont analysés ci-dessous. Le test de Levene a indiqué que ces valeurs avaient une distribution homogène.

- ***Prise de parole 'Pendant le menu'***

L'analyse de ce pourcentage a été réalisée dans les mêmes conditions que dans l'expérience 2. Le facteur *mode d'exploration* opposait le mode *'direct'* (présenté dans le Tableau 7-9, page précédente) au mode *'indirect'* (qui regroupe les modes *« balayage »* et *« picorage »* du Tableau 7-9. La procédure GLM intégrait le *nombre de tours de parole*, le *nombre de mots* et la *durée des dialogues* comme covariants.



La *'stratégie de présentation'* a eu un effet sur le taux de recouvrement :

***F(2,39)*=21,7 ; *p*<.00 ; η^2_p=.52**

Le mode d'exploration utilisé n'a pas eu d'effet :

F(1,39)=0,25 ; *p*=.61 ; η^2_p=.00

Et il n'y avait pas d'effet d'interaction :

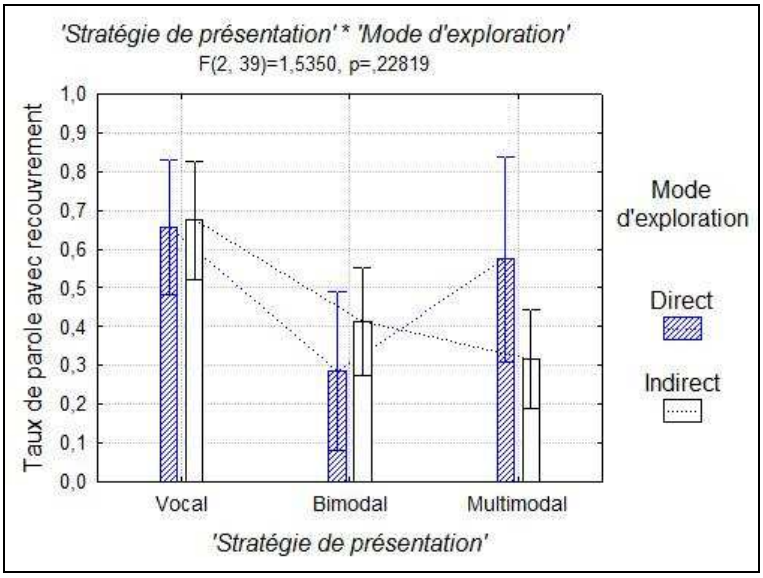
F(2,39)=1,55 ; *p*=.22 ; η^2_p=.07

Figure 7-13 : Taux de prises de parole *'pendant le menu'*

Les énoncés bimodaux et multimodaux ont conduits les participants à prendre la parole moins souvent pendant le menu que les énoncés vocaux.

- ***Taux de recouvrement entre parole de l'utilisateur et du système***

Le taux de recouvrement entre parole de l'utilisateur et parole du système porte sur la *phase de détail* dans le *dialogue simple*, comme cela a été présenté pour le moment de la prise de parole (ci-dessus). L'analyse réalisée intégrait les mêmes facteurs et les mêmes covariants que l'analyse du taux de prise de parole *'pendant le menu'*.



Le test de Levene a révélé une distribution homogène :

F(5,42)=1,73 ; *p*=.14

La *'stratégie de présentation'* a eu un effet sur le taux de recouvrement :

***F(2,39)*=22,5 ; *p*<.00 ; η^2_p=.53**

Le mode d'exploration utilisé n'a pas eu d'effet :

F(1,39)=1,03 ; *p*=.31 ; η^2_p=.02

Et il n'y avait pas d'effet d'interaction :

F(2,39)=1,53 ; *p*=.22 ; η^2_p=.07

Figure 7-14 : Taux de prises de parole avec recouvrement

On constate, comme c'était déjà le cas dans l'expérience 2, que la forme du graphique pour le taux de parole avec recouvrement est très proche de la forme obtenue pour le taux de parole *'pendant le menu'*.

- ***Vocabulaire : Taux d'utilisation des titres de films dans les commandes***

Le vocabulaire du service Cinéliste est relativement restreint. En dehors de la phase de requête en début de dialogue, les commandes les plus utilisées étaient : *« détail », « retour », « suivant », « précédent », « réécouter », « lecture »*. En plus de ces commandes, un vocabulaire un peu plus précis a parfois été utilisé pour naviguer dans le service. Il s'agissait de demandes directes d'accès à un film, soit en prononçant son numéro, soit en prononçant son titre. La Figure 7-15 porte sur les utilisations des titres des films dans les deux dialogues expérimentaux.

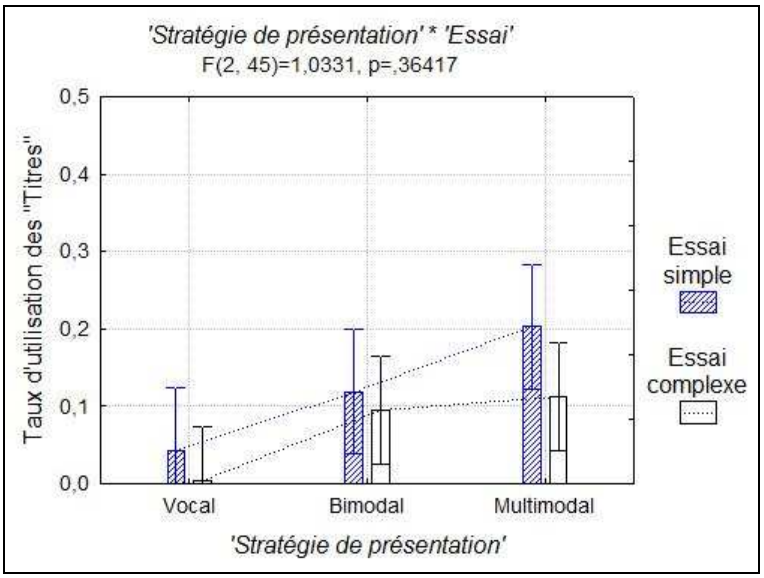


Figure 7-15 : Taux d'utilisation des titres de films

Les tests de Levene ont révélés que les distributions étaient déséquilibrées :
***F(2,45)*=9,69 ; *p*<.00**
***F(2,45)*=14,82 ; *p*<.00**

La *'stratégie de présentation'* a eu un effet sur l'utilisation des titres :
***F(2,45)*=4,22 ; *p*<.05 ; η^2_p=.15**

L'*essai* a également eu un effet :
***F(1,45)*=6,58 ; *p*<.05 ; η^2_p=.12**

Et il n'y avait pas d'effet d'interaction :
F(2,45)=1,03 ; *p*=.36 ; η^2_p=.04

On constate sur le graphique de la Figure 7-15 que les participants qui ont été confrontés à la *stratégie vocale* n'ont quasiment pas utilisé les *titres des films* dans leurs commandes alors qu'avec les deux autres stratégies plus de 10% des commandes incluaient un titre. Par ailleurs, ces commandes étaient plus fréquentes à l'essai simple qu'à l'essai complexe.

7.3.4 Conclusion

Du point de vue des indicateurs cognitifs cette expérience n'a pas permis de montrer un effet des stratégies de présentation testées. L'explication de ce résultat peut être recherchée entre autre dans la consigne donnée aux participants, comme cela a été indiqué dans la conclusion de l'expérience 2. La consigne de rappel explicite a induit la mise en place d'une stratégie comportementale spécifique qui a permis aux participants d'atteindre des performances équivalentes dans toutes les conditions, mais selon des procédures différentes. Le nombre de consultation des informations sur les films indique qu'un plus grand nombre de consultations était nécessaire avec la stratégie auditive. Le temps de lecture des informations indique une durée plus grande des consultations avec la stratégie bimodale. La stratégie multimodale est celle qui a conduit aux dialogues les plus naturels.

Du point de vue de la performance, il n'y a pas eu de gain de temps avec la *stratégie multimodale* sur la *stratégie bimodale*. Seule la *stratégie auditive* était moins performante puisque le temps de présentation des énoncés du système était plus long et puisque les participants ont consulté les informations un plus grand nombre de fois dans cette condition.

En revanche, le comportement stratégique était assez différent d'une stratégie de présentation à l'autre. Avec la *stratégie multimodale*, le nombre de consultation des informations sur les films était intermédiaire et le temps de consultation des informations était plus faible qu'avec la *stratégie bimodale*. De plus, les prises de parole des participants étaient plus fréquemment *'après la relance'*. Ces participants écoutaient plus facilement la totalité des énoncés du système, sans se précipiter. La variation conjuguée de ces indicateurs montre que le confort d'utilisation était plus important avec la *stratégie multimodale*.

L'exploration par *'balayage'* et l'utilisation des titres dans les commandes montrent une conjugaison entre le *mode d'exploration choisi* et le *vocabulaire utilisé* par les participants. Avec la *stratégie multimodale* les films étaient plus souvent désignés par leur titre. Les titres étaient utilisés dans les commandes car, quand ils étaient présentés visuellement, ils étaient lus. Avec la *stratégie auditive*, les *'réponses'* étaient, au contraire, sélectionnées au moment de leur présentation. Elles n'étaient pas réutilisées dans les commandes mais « attrapées au passage » à l'aide d'une commande unique (*« Détail »*). Dans ce cas, le processus consistait à glisser d'une réponse à l'autre par les commandes *« suivant »* / *« précédent »*, *« détail »* / *« retour »*. Ces deux modes d'exploration des données du service sont deux procédures distinctes qui relèvent du *processus collaboratif*, soit du *processus de construction des références*. La présentation visuelle des *références* (les *'réponses'*) a donné une présence à ces informations. Cette présence est devenue tangible dans le processus collaboratif parce que les participants ont engagé un comportement d'exploration des *'réponses'* visuelles tout en exerçant un contrôle sur le déroulement du dialogue dès que les *'échos'* ou les *'relances'* auditives venaient préempter cette activité. C'est ce phénomène qui a conduit les participants à lire les titres de films. Pour l'emploi de la reconnaissance vocale, ce type de résultat indique qu'il est possible de « forcer » l'emploi de certains termes en utilisant ce type de stratégie de présentation bimodale ou multimodale. Dans cet exemple, l'emploi des titres de films est préférable pour la reconnaissance vocale que l'emploi de numéros, qui peut entraîner des substitutions (par exemple entre *« film six »* et *« film dix »*).

Chapitre 8 Discussion

Les objectifs de la thèse portaient sur l'*analyse des énoncés des systèmes de DHM* pour tenter d'en déduire des principes de conception utilisables dans le cadre du développement de ces systèmes. L'enjeu des expériences présentées était de proposer et d'évaluer des *'stratégies de présentation des informations'* basées sur certaines règles d'analyse dans le but de valider les principes de conception issus de ces règles d'analyse. La validation expérimentale de ces principes doit permettre d'envisager la conception de systèmes de DHM plus compétents dans la communication, *i.e.* susceptibles d'adapter leur comportement dans l'interaction pour produire la diversité des effets souhaitables en vue de la réussite de la communication. Ces questionnements renvoient aux causalités internes au processus de communication et ils permettent d'aborder des points théoriques fondamentaux.

Cette partie propose d'abord une synthèse des résultats obtenus dans les cinq expériences de la thèse. Les implications de ces résultats, tant pratiques que théoriques sont abordées ensuite. Les limites et les perspectives sont envisagées dans la dernière partie.

8.1 Synthèse des résultats

8.1.1 Dialogue Homme Machine vocal

Les expériences 1 et 2 portaient sur le DHM vocal. L'objectif était de vérifier la nécessité de certaines informations fréquemment présentées dans le cadre du fonctionnement normal de systèmes de DHM (systèmes commerciaux). Les messages d'aide donnent lieu à de longues explications. De même, l'utilisation d'une syntaxe complète donne lieu à une verbosité importante du système dont on se demande si elle est excessive. Il s'agissait d'évaluer les effets que produisent ces informations.

A Expérience 1

L'expérience 1 a permis de montrer que les messages d'aide n'entraînaient pas en eux-mêmes de gêne pour la mémorisation des informations recherchées. Dans le contexte du DHM, ces messages n'ont pas entrainé directement un processus de compétition pour les ressources. Les participants ont mis en place un mécanisme inférentiel intermédiaire (lié à la motivation) qui les a conduits à exercer un certain mode de contrôle sur l'activité et à investir certaines ressources. C'est l'utilisation d'un mode de contrôle ou d'un autre qui a entraîné des effets sur le rappel et sur la charge cognitive subjective. Les meilleures performances ont été

obtenues par les participants qui ont exercé le meilleur contrôle. Les participants qui ont exercé un contrôle plus passif ont obtenu des performances moyennes ou faibles.

Le rôle des aides a consisté à indiquer aux participants quels automatismes ils pouvaient utiliser dans les procédures de contrôle du dialogue. Les effets étaient comportementaux. Ils ont montré que les *'aides réparties'* étaient préférables parce qu'elles permettaient un contrôle plus souple, sur un mode conversationnel, selon les principes du modèle sociocognitif (Clark, 1996) et du modèle de l'alignement interactif (Pickering & Garrod, 2004).

B Expérience 2

Dans l'expérience 2, la suppression de la syntaxe n'a pas entrainé de modification de la performance ni de l'effort investi par le participant. Cependant, quand le système avait une verbosité moins importante, les participants ont réécouté les informations un plus grand nombre de fois, ce qui indique soit une difficulté de compréhension, soit au contraire, une certaine facilité d'écoute.

Dans la condition de '*style espacé'*, plus lente, la fréquence importante des commandes en un seul mot montre que ces énoncés ont incité les participants à s'aligner sur ce style *'mot isolé'*, conformément à l'effet de durée de la parole (Matarazzo et al., 1963; Zoltan-Ford, 1991). Dans la condition de *'style normal'*, plus verbeuse, les commandes des participants ont été moins souvent en un seul mot. Les participants étaient également plus verbeux. Dans cette condition, les taux de recouvrement élevés indiquent qu'un phénomène de lutte pour la prise de parole a occupé ces participants pendant les dialogues.

Ces résultats indiquent d'abord, que la performance dans le dialogue dépend de la procédure mise en place par le participant (expérience 1) et ensuite, que certains aspects de cette procédure dépendent de la forme des comportements du système (structure syntaxique des énoncés, expérience 2). Mais sur ce second point, la manipulation de la syntaxe des énoncés a produit des effets faibles. Après ces résultats, il était supposé que l'utilisation de la modalité visuelle pouvait produire des effets plus conséquents.

8.1.2 Dialogue Homme Machine multimodal

Les expériences 3, 4 et 5 portaient sur l'utilisation complémentaire de la modalité visuelle pour présenter les informations verbales habituellement présentées sur la modalité auditive dans les systèmes de DHM. L'ajout de la modalité visuelle était supposé réduire les contraintes liées au traitement des informations présentées à l'utilisateur.

A Expérience 3

L'expérience 3 était basée sur l'assimilation de la situation de présentation des informations en DHM à une situation de présentation d'une liste d'items suivie d'un suffixe non pertinent dans un test de rappel. Elle a permis de montrer que l'*effet de suffixe* est présent, dans le

cadre du DHM, quand les *'réponses'* du système sont présentées sur la modalité auditive. Cet effet est d'autant plus important si les *'relances'* sont présentées sur la même modalité.

Pour faire cette comparaison, la catégorisation des informations présentées par le système a été nécessaire. L'assimilation des *'réponses'* à la liste des items à mémoriser a conduit à assimiler les *'relances'* au suffixe « non pertinent ». Or, un effet spécifique à ce type d'information a également été identifié. Quand les *'relances'* étaient présentées auditivement (conditions *'auditive'* et *'redondante'*) les participants répondaient plus vite. Cet effet était distinct de l'effet du mode de présentation des *'réponses'*. Quand les *'réponses'* étaient présentées visuellement (condition *'visuelle'* et *'redondante'*) les participants répondaient plus lentement. Ces effets étaient en interaction dans l'une des phases du dialogue (phase de *sélection*) mais pas dans l'autre (phase d'*information*). Ainsi, cette expérience a permis de montrer que *la modalité* utilisée pour présenter une information produit des effets différents en fonction du *type d'information,* catégorisé sur une base fonctionnelle.

B Expérience 4

L'expérience 4 a permis d'étudier plus spécifiquement cette *'relation type-mode'*. Les trois types d'information présents dans tout dialogue homme machine (*'écho'*, *'réponse'*, *'relance'*) ont été associés préférentiellement à l'une des deux principales fonctions dialogiques (*'information'* ou *'contrôle'*) pour proposer quatre *stratégies de présentation des informations*.

Les résultats ont montré que les deux stratégies bimodales (AVA et VAV) n'étaient pas équivalentes. L'une de ces stratégies (VAV : *contrôle visuel – information auditive*) a conduit au rappel le plus faible. Les résultats de l'expérience 3 sur l'effet de suffixe permettaient pourtant d'attendre un rappel plus faible avec la stratégie entièrement auditive (AAA). Les résultats indiquent (1) qu'avec cette stratégie bimodale (VAV) les participants ont tenté d'exercé le contrôle du dialogue pendant la présentation des *'réponses'*, (2) qu'ils ont moins pris en compte les *'échos'* et les *'relances'*, et (3) qu'ils ont été plus verbeux face au système. Cette fois encore, le mode de contrôle de l'activité est une variable intermédiaire importante. L'utilisation conjointe des modalités visuelle et auditive n'est pas suffisante en elle-même pour accroître la performance des participants.

La stratégie AVA (*contrôle auditif – information visuelle*) a été celle qui donnait lieu à la meilleure performance : (1) le rappel était aussi bon dans cette condition que dans la condition entièrement visuelle (stratégie VVV), (2) le gain de temps par rapport à la stratégie auditive (AAA) était de 35%, (3) les prises de parole apparaissaient après la *'relance'* et (4) le nombre de mots par tour de parole était équilibré d'une phase du dialogue à l'autre. Cette stratégie a donné lieu au contrôle le plus naturel. Le gain de temps par rapport à la stratégie auditive (AAA) était inférieur au gain de temps avec la stratégie VVV, de 45%. Mais avec cette stratégie (VVV) les participants ont été particulièrement verbeux, ce qui est négatif pour l'usage de la reconnaissance vocale. La stratégie AVA n'a pas provoqué cet effet. Ainsi, cette expérience permet de valider l'utilité de la prise en compte de la *relation type-mode*, identifiée dans l'expérience 3, pour l'opération d'attribution des modes de présentation.

C Expérience 5

L'expérience 5 avait pour but de vérifier si des gains de performance plus importants peuvent être obtenus, grâce à des formats de présentation bimodaux, dans un dialogue présentant des données plus riches. La *'stratégie entièrement auditive'* (AAA) a été comparée à la *'stratégie bimodale'* la plus favorable (AVA) et à une *'stratégie multimodale'* conçue sur les bases d'une analyse plus riche des contenus à présenter.

Toutes les stratégies de présentation utilisées dans cette expérience ont permis aux participants d'atteindre leurs objectifs, qui consistaient à explorer des données riches pour sélectionner une cible (un film) et à mémoriser les informations se rapportant à cette cible. Les deux stratégies bimodales (*'bimodale simple'* et *'multimodale'*) ont permis des gains de temps (de l'ordre de 50% et de 45%, respectivement).

Les résultats n'ont pas permis de mettre en évidence un effet direct sur le rappel ni sur la charge cognitive. Cependant, l'une des stratégies (la *stratégie entièrement auditive*) a conduit les participants à réécouter les informations un plus grand nombre de fois ; et une autre (la *stratégie bimodale simple*) les a conduit à de longs temps de silence pendant la lecture (les moyennes allaient jusqu'à 30 secondes). La *stratégie multimodale enrichie* était préférable car elle permettait d'éviter ces deux écueils. Cela indique que les principes de conception utilisés pour proposer cette *stratégie de présentation* ont permis une interaction plus fluide.

8.1.3 Evaluations subjectives de la charge cognitive

En ce qui concerne les évaluations subjectives de charge cognitive, les expériences 1 et 4 confrontaient deux questionnaires concurrents : *'NASA-TLX'* et *'Workload Profile'*. Dans l'expérience 3, c'est le premier qui a été utilisé (*'NASA-TLX'*). Dans les expériences 2 et 5, c'est le second qui a été utilisé (*'Workload Profile'*). Ces deux questionnaires se sont révélés sensibles à certaines des variables expérimentales, mais ils ont mis en évidence des variations de natures différentes.

Le questionnaire *'NASA-TLX'*, basé sur une décomposition des sources de coût liées à la tâche, a permis de mettre en évidence deux effets dans les expériences 1 et 3, mais pas dans l'expérience 4. Dans l'expérience 1, ces évaluations ont été sensibles à un effet d'ordre. D'après cet effet, les participants ont trouvé le second dialogue plus facile que le premier, indépendamment des facteurs expérimentaux. Dans l'expérience 3, ces évaluations ont révélé l'effet du mode de *'réponse'*, conformément aux résultats obtenus pour le rappel. 'NASA-TLX' a révélé que la tâche était considérée comme plus difficile quand les *'réponses'* n'étaient présentées que sur la modalité auditive. En revanche, dans l'expérience 4, ce questionnaire n'a permis de mettre en évidence aucun effet alors que les évaluations obtenues avec *'Workload Profile'* étaient sensibles à plusieurs facteurs.

Le questionnaire *'Workload Profile'*, basé sur une décomposition des pools de ressources individuels, a permis de mettre en évidence des effets dans les expériences 1, 2 et 4, mais

pas dans l'expérience 5. Dans l'expérience 1, ces évaluations ont été sensibles à la *position du dialogue avec erreurs* de reconnaissance vocale et à un léger effet d'interaction avec la *condition d'aide*. Elles ont révélé une plus grande sensibilité des participants à l'apparition précoce d'erreurs de reconnaissance dans la condition *'aide réparties'*. Dans l'expérience 2, ces évaluations ont révélé l'effet simple de la consigne. Les dialogues complexes ont été évalués comme plus difficiles. Dans l'expérience 4, ces évaluations ont été sensibles à la présence d'une erreur de reconnaissance vocale, et les comparaisons planifiées ont révélé que cette sensibilité était spécifique à la stratégie de présentation entièrement auditive (AAA). De plus, ces évaluations ont été très sensibles au mode de correction de l'erreur utilisé par les participants, ce qui pointe le rôle essentiel des facteurs comportementaux. En revanche, dans l'expérience 5, ce questionnaire n'a révélé aucun effet des facteurs expérimentaux. Mais dans cette expérience, les autres indicateurs généraux ont également été peu affectés.

'Workload Profile' a été le questionnaire le plus sensible aux facteurs expérimentaux. Ce questionnaire a fourni les résultats les plus intéressants pour l'analyse. Les *analyses discriminantes*, qui prennent en compte l'ensemble des dimensions du questionnaire pour catégoriser les groupes expérimentaux, ont permis de révéler des *profils de charge cognitive* différents entre les différentes conditions de l'expérience 4. Ce résultat indique que les stratégies de présentation ont bien eu un effet sur les pools de ressources que les participants ont dû activer pour mener à bien les tâches qui leur étaient prescrites. Mais la réponse comportementale reste l'intermédiaire nécessaire. Notamment, le détachement du groupe de participants ayant utilisé la stratégie entièrement visuelle (VVV) vers la réponse manuelle (*Cf.* Figure 7-9, p. 194) a permis de comprendre qu'avec cette stratégie, les participants ont eu un mouvement d'alternance plus difficile entre le planning qui leur donnait leurs contraintes (fiche remise au participant au début de l'expérience) et les propositions du système (simulation sur ordinateur). D'après ce résultat, *'Workload Profile'* peut être considéré comme un questionnaire ayant une bonne valeur diagnostique, comme le précisent Tsang et Velasquez (1996) dans leur article de présentation de cet instrument.

'NASA-TLX' a également eu une sensibilité à certains facteurs, mais il n'a mené à aucun raisonnement de recherche de cause et on ne peut considérer que ce questionnaire ait eu une valeur diagnostique dans le cadre des tâches étudiées. Cette comparaison des deux questionnaires dans des tâches de DHM est à l'avantage de *'Workload Profile'*.

8.1.4 Conclusion

Ces résultats indiquent d'abord que, dans le DHM, il est plus facile d'obtenir des gains d'efficacité en utilisant la modalité visuelle qu'en manipulant uniquement la forme et le contenu vocal des énoncés. Les gains temporels sont importants dès que la modalité visuelle est utilisée (jusqu'à 50% de temps gagné en moyenne). Mais plus généralement, c'est toute la structure de l'interaction qui est modifiée. Dans une certaine mesure, l'utilisation de la modalité visuelle permet d'orienter l'attention et le comportement de l'utilisateur sans qu'il soit

nécessaire de passer par des consignes explicites comme c'est le cas dans les messages d'aide. Dans l'interaction, les modalités visuelle et auditive ont d'autres fonctions que celle de *'présenter des informations'*.

Ces résultats indiquent que toutes les parties d'un énoncé ne sont pas équivalentes entre elles et que la règle d'attribution des *modes de présentation* aux différentes *unités informationnelles* n'est pas indifférente. Une catégorisation d'ordre fonctionnel (*'écho'*, *'réponse'*, *'relance'*) a été introduite pour différencier les contenus qui composent les énoncés du système. Selon cette catégorisation, l'énoncé n'est pas vu comme une liste d'informations à acquérir (en mémoire), mais comme une série d'actes destinés à produire des effets sur les interlocuteurs et dans le dialogue relativement à des conventions partagées par les partenaires. Grâce à cette catégorisation, les énoncés du système ont été découpés en éléments discrets et des règles d'attribution des modes de présentation à ces éléments ont permis de concevoir des *'stratégies de présentation'* pour en tester les effets dans les dialogues expérimentaux. Les expériences proposées ont permis de mettre en évidence une diversité d'effets, qui portent :

- *Sur les automatismes* mis en place par les participants dans les dialogues (prises et cessions de parole, dans les cinq expériences de la thèse) ;
- *Sur les modes de contrôle*, avec l'orientation des participants vers certaines procédures pour la correction des erreurs (expériences 1 et 4) et pour l'exploration des réponses (expériences 2 et 5) ;
- *Sur la réalisation de la tâche*, avec des modifications de la durée des dialogues (surtout en mode visuel : expériences 3, 4 et 5), du nombre de mots et du nombre de tours de parole (expériences 1, 2, 4 et 5) ;
- *Sur la performance cognitive* résultante, avec le rappel et les évaluations subjectives de charge cognitive (expériences 1, 3 et 4).

Ces résultats permettent de valider l'intérêt de la *catégorisation* et des *règles* utilisées pour la conception des énoncés, car elles ont permis d'obtenir une interaction plus souple et plus efficace. Ils indiquent qu'une conception additive des ressources cognitive consacrées aux différentes sous-activités est insuffisante pour prédire la richesse des effets produits dans un processus multifonctionnel tel que le dialogue et sur un individu doué de capacités d'inférence. La décomposition analytique (élémentariste et fonctionnelle) de la situation de communication est nécessaire pour comprendre ces effets. Les principes de l'analyse pragmatique ont permis de faire cette analyse et d'expliquer les résultats obtenus.

8.2 Implications

Des implications de plusieurs types peuvent être dégagées à partir de ces résultats. Les premières abordées portent sur la conception des énoncés dans le DHM. Un point permet ensuite d'aborder le problème de la conception des énoncés multimédias en contexte d'apprentissage. Les conséquences pour l'utilisation des concepts théoriques utilisés pour l'explication sont ensuite abordées. Elles portent sur le statut accordé aux notions de *charge cognitive*, d'*information* et d'*action*. Enfin, le dernier point permet de revenir sur la notion de *pertinence* en pragmatique.

8.2.1 *'Stratégies de présentation'* en DHM

La démarche de conception des énoncés utilisée dans le cadre des expériences 4 et 5 de la thèse pourrait être utilisée pour la conception de systèmes dialogiques destinés au grand public. Les gains de temps et la souplesse de l'interaction avec la *stratégie AVA* (*information visuelle – contrôle auditif*) dans l'expérience 4 et avec la *stratégie multimodale* dans l'expérience 5 permettent d'envisager des systèmes à la fois plus efficaces et plus conviviaux.

A Pour l'implémentation des 'stratégies de présentation'

La conception de systèmes dialogiques qui utiliseraient ces *'stratégies de présentation'* suppose cependant de s'en donner les moyens techniques et de développer des plates-formes de communication et des interfaces capables d'adopter ces comportements. Dans le contexte d'une interaction téléphonique, ces principes supposent l'existence de dispositifs de diffusion en parallèle sur les deux modalités principales (visuelle et auditive) et de moyens de contrôle de ces dispositifs à un niveau central du système (module de *gestion du dialogue*, *Cf.* Figure 2-2, p. 43). Des terminaux tels que les iPod® supporteraient facilement ce type de techniques. Mais ces extensions impliquent également d'adapter le fonctionnement des serveurs qui diffusent les informations[1] à travers les réseaux. D'une part, des techniques spécifiques doivent être utilisées pour formaliser les connaissances nécessaires dans le système et, d'autre part, le processus de développement dans son ensemble doit être adapté (à chacune des étapes du cycle itératif : *spécification, développement, tests*). De plus, le développement de systèmes adaptatifs[2] suppose de faire appel aux techniques de l'*intelligence artificielle*. Grâce à ces techniques, il est possible de formaliser les principes d'analyse proposés dans la thèse. En l'absence d'une telle formalisation, il est possible de faire adopter des patterns comportementaux prédéfinis au système (comme dans le protocole

[1] Ici, le terme *« information »* a une signification technique.

[2] Les systèmes dits *« adaptatifs »* sont des systèmes capables d'adaptations autonomes au cours d'une activité.

en *Magicien d'Oz* utilisé dans la thèse). Dans ce cas, l'analyse des fonctions dialogiques est faite par les concepteurs lors du développement du système. Elle est convertie sous la forme d'une séquence comportementale et le système est développé pour produire cette séquence (sans en faire l'analyse). Dans ce cas, le système peut être dit « adapté », mais il ne s'agit pas d'un *« système adaptatif »*. Pour faire adopter ce type de comportement à des systèmes, des évolutions des plates-formes téléphoniques classiques sont cependant nécessaires. Le standard de programmation VoiceXML[1] indique par exemple qu'il est nécessaire de pouvoir prendre en compte ce type de comportements. Mais les principes de conception des stratégies de présentation proposées dans les expériences 3 et 4 n'ont aujourd'hui été identifiés (par l'auteur de la thèse) dans aucun service commercial qui exploiterait ces possibilités, ni dans aucune étude publiée.

Comme cela a été indiqué au chapitre 2, la *théorie de l'interaction rationnelle* (Sadek, Bretier, & Panaget, 1997) est aux confins de ces aspects puisqu'elle a permis d'associer les techniques de formalisation de l'intelligence artificielle aux techniques d'interaction téléphonique. Dans ce cadre, les travaux de Clémente (2004) ont permis de formaliser les connaissances du système sur les dispositifs techniques dont il dispose à un moment donné pour présenter des informations. A sa suite, l'approche d'Horchani (2007) a permis d'initier un travail de formalisation qui tentait d'associer les choix relevant de la *'stratégie de présentation'* et de la *'stratégie de dialogue'* (deux niveaux de décision généralement traités dans des modules différents dans les systèmes dialogiques). Ces travaux ont donné lieu à une tentative de formalisation des stratégies de présentation utilisées dans l'expérience 4. Un outil de spécification des *stratégies de présentation* a été proposé à l'issue de ce travail (voir Horchani, Fréard, Caron et al., 2007; Horchani, Fréard, Jamet, Nigay, & Panaget, 2007a; Horchani, Fréard et al., 2007b) et devait être utilisé pour le développement du matériel des expériences 2 et 5. Malheureusement, des dysfonctionnements techniques ont imposé de se rabattre sur les outils utilisés dans les expériences précédentes (développement de scripts à l'aide de Macromedia Director®).

De façon plus fondamentale, ces travaux étaient dédiés à la conception d'un outil de développement des stratégies de dialogue multimodal (Horchani, Nigay & Panaget, 2007). Mais ils n'abordaient pas la question des effets des actes et de l'analyse fonctionnelle nécessaire à la conception d'une règle d'attribution des modes de présentation aux différentes unités de contenu. Il s'agissait plutôt, dans ce cas, de prédéfinir des patterns comportementaux réutilisables dans différents contextes, que d'implémenter la dynamique des raisonnements sur les variations contextuelles. La conception d'un algorithme qui serait dédié à cette règle d'attribution suppose de formaliser (dans le système qui supporte cet algorithme) les connaissances nécessaires à cette analyse fonctionnelle. Les travaux de Keizer et Bunt (2006 ; 2007a ; 2007b), présentés en fin de chapitre 2, offrent des perspectives intéressantes dans ce sens. En complément à ces approches, les travaux de la

[1] Par exemple, les recommandations présentées par le W3C (http://www.w3.org/TR/vxml30reqs/) intègrent ces possibilités (voir paragraphe : http://www.w3.org/TR/vxml30reqs/#mod-csmo).

thèse ont permis d'illustrer les arguments de Le Bigot et al. (2007) en faveur de l'utilisation d'une diversité d'indicateurs pour analyser le processus collaboratif en dialogue. Des systèmes adaptatifs doivent disposer de moyens leur permettant d'analyser les différents effets en intégrant tous les indicateurs de façon à prendre les bonnes décisions d'action. Cela suppose que différents niveaux d'interprétation doivent être pris en compte en parallèle (voir Clark, 2002 ; Allwood, 1995) et qu'un dispositif de contrôle central doit être en mesure de mettre ces connaissances en *résonance* (selon une métaphore vibratoire – voir plus loin –) pour identifier le meilleur cheminement dans la *chaîne causale*. Tout un ensemble de connaissances est nécessaire pour aller dans ce sens. L'implémentation de systèmes intégrant ce type de formalisation permettrait d'éprouver et d'approfondir cette analyse pour enrichir la description des relations causales internes au processus de communication.

Pour aller plus loin, on peut noter que les niveaux de coordination proposés par Clark (2002) et par Allwood (1995) se rapportent à une décomposition hiérarchique des interactions (*Cf.* Roulet et al., 1985. Voir Tableau 1-6, p. 18). Les trois plans d'intentions qui ont été distingués dans la Figure 5-1 (page 124) renvoient à trois types de *connaissances relationnelles* (spatio-temporelles, sémiotiques et sociales) qui peuvent être croisées à ces niveaux d'observation du comportement de l'utilisateur. La prise en compte des intentions de l'utilisateur dans ces trois plans pour chaque séquence d'action qu'il produit devrait permettre d'envisager des systèmes capables d'interprétations riches sur les finalités de l'action en cours, pour en dégager la signification. Cela suppose que le système dispose de bases de connaissances à la fois sur (1) les relations sociales, (2) les relations sémiotiques et (3) les relations spatio-temporelles, ainsi que (4) d'un mécanisme capable de mettre en œuvre ces connaissances et de les utiliser pour interpréter les actes. La mise en place de ce type de système pourrait permettre d'envisager des architectures cognitives différentes de celles proposées dans les approches américaines évoquées au chapitre 4 (notamment ACT-R et EPIC). Un système mettant en place ce type d'interprétation et capable de mettre en œuvre des actions correspondantes devrait être capable d'offrir des adaptations comportementales complexes basées sur des raisonnements causaux finalistes. Dans un tel système, des finalités diverses peuvent être visées et hiérarchisées pour rechercher, par exemple, l'accroissement des connaissances de l'utilisateur dans un domaine particulier ou l'optimisation de ses compétences dans l'utilisation d'un outil en lui expliquant des procédures qu'il exécuterait de façon incomplète ou fallacieuse. Evidemment, un travail de convergence entre des techniques diverses est encore nécessaire. Mais l'émergence actuelle de la *robotique humanoïde* permet de rêver à des perspectives lointaines[1].

B Formalisation des actes dans les systèmes adaptatifs

L'objectif de l'analyse proposée est de donner à des systèmes dialogiques des capacités de raisonnement sur leurs actes sur un mode abstrait. Ces systèmes pourraient ainsi devenir

[1] Voir par exemple le site de la société Aldébaran : http://www.aldebaran-robotics.com/ ou les pages sur le projet FLOWERS de l'INRIA : http://www.inria.fr/recherche/equipes/flowers.fr.html

capables d'exercer une influence positive sur le comportement de l'utilisateur (voir par exemple, Zoltan-Ford, 1991 ; Bubb-Lewis &Scerbo, 2002). Pour cela, ils doivent être capables de produire des actes qui causeront les effets souhaités. Cette production repose sur deux mécanismes élémentaires :

1. Un *mécanisme de catégorisation des actes*, qui suppose de disposer de connaissances préalables sur les catégories existantes et d'une possibilité d'associer les *unités de contenu* à présenter à ces connaissances prototypiques. Cela suppose que des rôles prototypiques peuvent être associés à ces actes. Les *actes prototypiques* pris en compte dans la thèse étaient des *'échos'*, des *'réponses'* et des *'relances'* (*Cf.* chapitre 2. pp. 54-55). Des catégorisations plus fines sont envisageables, notamment pour accorder un statut spécifique aux messages d'aide (à ce sujet, voir Babin, 2007) ;
2. Un *mécanisme d'attribution des modes de communication* aux différents types d'actes doit permettre de décider comment incarner ces actes pour qu'ils remplissent leurs rôles au mieux (*i.e.* qu'ils atteignent les effets visés). Par exemple, la présentation de l'*écho'* sur la modalité auditive implique qu'un contrôle immédiat est attendu (la mise en place immédiate d'une procédure de correction est attendue en cas d'erreur) alors que la présentation de cet *écho'* sur la modalité visuelle implique que le contrôle de cette information est moins urgent ou qu'il pourrait demeurer nécessaire au cours du temps. Des implications de ce type peuvent être associées à chaque acte.

L'attribution des modes de communication aux actes qui composent un énoncé dépend de la *relation type-mode*. Cette relation indique que les différents modes de communication qui peuvent être identifiés (ici, forme auditive ou visuelle) réalisent les fonctions dialogiques avec une efficacité différente. Cette efficacité est relative aux trois plans d'intentions qui ont été identifiés. Elle dépend des propriétés physiques de l'organisme à la source de l'acte (ici, le système) et de l'organisme qui le perçoit (l'utilisateur). Le Tableau 8-1 (page suivante) est une proposition pour tenter de synthétiser cette relation.

Dans la *théorie de l'interaction* (Sadek, Bretier, & Panaget, 1997), la formalisation des actes ne permet de représenter explicitement que la *fonction d'information*. Les noms donnés aux actes permettent de l'illustrer (INFORM, INFORMSI, DEMANDE, etc.) Les deux autres types de fonctions ne sont pas explicitement pris en compte. Le fait de représenter les relations *sociales* et *spatio-temporelles* dans un formalisme de ce type (*Cf.* chapitre 2, pp. 45-48) permettrait d'engager le système dans des raisonnements prenant en compte d'autres aspects que le contenu asserté en lui-même. Par exemple, dans le *plan interactif* (relations spatio-temporelles) les temps de réaction de l'utilisateur peuvent être analysés en les rapportant aux actes auxquels ils répondent, la vitesse d'élocution peut être interprétée, ainsi que le taux de recouvrement, etc. Ces données sont souvent prises en compte dans les systèmes, mais elles ne sont pas analysées sur la base de leurs finalités. Pour cela, la diversité des indicateurs pris en compte est importante et le modèle conceptuel qui permet de les intégrer est essentiel. Ces remarques indiquent qu'il est nécessaire de disposer de connaissances sur les actes, de façon à pouvoir les traiter à un niveau abstrait.

Tableau 8-1 : Principes pour la *prédiction des effets des actes* dans l'interaction

Principe de ***distinction fonctionnelle***	Tout acte rempli trois fonctions générales dans la communication : (1) La ***fonction de communication*** vise à produire une situation d'échange entre les partenaires (*affirmation, question, remerciement, promesse, etc.*). Elle renvoie, par exemple, aux *obligations de réponses* ; (2) La ***fonction d'information*** vise à faire connaître un ensemble de propositions (*le contenu asserté*) ; (3) La ***fonction d'interaction*** vise à provoquer chez les destinataires des comportements en réaction à ces propositions (*temps de réaction, vitesse d'élocution, prosodie, mouvement, etc.*) Ces fonctions correspondent à des faisceaux d'intentions basés sur des réseaux de relation distincts.
Principe de ***recouvrement fonctionnel***	Les fonctions de *communication*, d'*information* et d'*interaction* sont susceptibles d'être réalisées par le biais de différents modes de présentation (*e.g. vocal* ou *écrit*). Chaque boucle d'action (1. *vision-geste*, 2. *vision-parole*, 3. *audition-geste*, et 4. *audition-parole*) fonctionne selon des contraintes qui lui sont propres. L'utilisation d'un mode de présentation ou d'un autre pour présenter un acte permet à cet acte d'engendrer les effets attendus avec une réussite plus ou moins importante relativement à chacune des fonctions identifiées. Les effets des actions engagées doivent être prédits pour formuler des hypothèses qui peuvent être vérifiées dans la suite de l'interaction sur la base des actions ultérieures de l'utilisateur.

Ces propositions d'extension des descriptions des actes en classant leurs effets sur trois plans permettent d'imaginer des systèmes capables de faire des inférences plus fines (*i.e.* susceptibles de produire de la signification) pour s'adapter aux utilisateurs des services.

C Conséquences pour la modélisation de l'utilisateur

Dans les systèmes de dialogue, le modèle de l'utilisateur est généralement conçu à partir du modèle de la langue. De façon schématique, le modèle de l'utilisateur est la combinaison formée par l'ensemble des phrases susceptibles d'être produites par l'utilisateur (McTear, 2002). Traduite dans le langage des commandes du système, cette combinaison de phrases correspond à l'ensemble des commandes qui peuvent être produites auprès du système. Autrement dit, pour le système, le modèle de l'utilisateur est l'ensemble des actions de cet utilisateur qu'il est susceptible de prendre en compte. Des principes de calcul de plan peuvent être utilisés pour prédire les chaînes d'actions qui apparaissent le plus fréquemment dans des séries de commandes liées à l'exécution d'une tâche pour donner des compétences adaptatives au système (Louis, 2002 ; Baudoin, 2007).

Les travaux de Brennan et Hulteen (1995) ont permis de montrer comment un système peut adapter ses niveaux de feedback à l'utilisateur en se basant sur l'historique du dialogue. Sur ce principe, la prise en compte simultanée des plans *interactif*, *assertif* et *communicatif* pour qualifier les actes des utilisateurs pourrait permettre d'identifier des variations et des constantes intra et interindividuelles qui pourraient permettre des adaptations plus fines. Certains individus pourraient être regroupés en fonction de « styles » qui peuvent apparaître

dans leurs modes de communication. La possibilité d'identifier ces styles peut permettre d'y adapter le comportement du système selon des principes de mimétisme dont ont sait qu'ils sont utiles à la communication (*e.g.* Oviatt, Darves & Coulston, 2004). Des *préférences* individuelles (sous forme de *patterns comportementaux préférentiels*) pourraient apparaître relativement à la règle d'attribution des modes de communication. Par ailleurs, l'identification de ces styles peut permettre de définir un système d'attentes sur le comportement de l'utilisateur au cours de l'interaction, qui pourrait permettre de diagnostiquer les échecs et d'interpréter les réussites dans la tâche. – Pour la formalisation d'un système d'attentes (*« expectations »*), voir Castelfranchi (2007). – Enfin, la définition de ces styles pourrait permettre de simuler des styles de personnalités différents au travers de styles comportementaux différents.

De tels travaux autour de systèmes capables d'interprétations dynamiques (voir Bunt, 2000) peuvent apporter des éléments intéressants, par exemple, pour l'étude de la personnalité.

8.2.2 Etude des énoncés en situation d'apprentissage

Comme cela a été expliqué dans la présentation des théories de l'apprentissage multimédia (*Cf.* chapitre 3), ces approches relèvent d'une problématique de conception (*« design »*) bien que les auteurs focalisent leur attention sur le processus d'apprentissage. Des critiques formulées dans la partie théorique permettaient de préciser certaines des limites de la synthèse théorique des auteurs (notamment, Mayer, 2005 ; Moreno & Mayer, 2007). D'une part, ces auteurs se focalisent excessivement sur la tâche d'apprentissage et tendent à laisser de coté l'analyse des autres aspects de l'activité (qui coexistent avec l'apprentissage pendant l'apprentissage). D'autre part, la nature conventionnelle de la relation pédagogue-apprenti (dimension illocutoire) n'est pas questionnée par ces auteurs. D'après leur analyse, dans la situation d'apprentissage, des informations circulent du pédagogue vers l'apprenti, sans autre procès. Sur ce principe, les modalités perceptives sont vues comme des canaux d'entrée à capacité limitée. Cela conduit notamment à expliquer l'*effet de modalité* par l'additivité des ressources visuelles et auditives (*e.g.* Ginns, 2005 ; Tindall-Ford, Chandler & Sweller, 1997).

La chaîne causale proposée peut être représentée de la manière suivante :

Format de présentation => Ressources cognitives => Apprentissage

Dans cette chaîne, les principes de conception des énoncés ont un effet sur la quantité de ressources disponibles, ce qui produit un effet sur la performance d'apprentissage. La théorie de la charge cognitive repose sur cette relation causale. Elle consiste à affirmer cette relation. Cependant, cette relation est aujourd'hui modérée par les auteurs. Notamment, la méta-analyse de Ginns (2005) montrait que l'effet de modalité doit être modéré par l'*interactivité du système* et par la possibilité d'un *contrôle pas-à-pas* sur le défilement des informations. De même, Gerjets et Scheiter (2003) indiquaient que les *buts* et l'*activité* de l'enseignant ont un effet sur les patterns de charge cognitive de l'individu apprenant et sur la performance

d'apprentissage qui en résulte[1]. Pour aller plus loin, il serait nécessaire de remettre en question la relation de causalité que ces auteurs proposent de modérer. Les résultats de la thèse indiquent qu'il est nécessaire de faire cette remise en cause.

L'*hypothèse d'additivité des ressources* laissait supposer que les stratégies de présentation utilisant les modes de communication de façon complémentaire (fission sur deux canaux, sans redondance) permettraient de réduire la charge cognitive des participants et, par suite, d'accroître leur performance d'apprentissage. Dans les expériences de la thèse, toutes les informations à présenter aux participants étaient verbales, ce qui a permis d'opposer les modes de présentation pour proposer les différentes stratégies. Plusieurs *stratégies bimodales* concurrentes pouvaient être comparées (stratégies AVA et VAV dans l'expérience 4, stratégies AVA et multimodale dans l'expérience 5). Dans ce contexte, l'hypothèse théorique d'une additivité des ressources conduisait à l'hypothèse opérationnelle d'un *effet de bi-modalité* (nommé ainsi pour le différencier de l'*effet de modalité* classique qui associe image et texte), qui stipule que toutes les stratégies bimodales devaient permettre un meilleur apprentissage des contenus. L'expérience 4 a été conçue pour tester cet *effet de bi-modalité* et éprouver l'*hypothèse d'additivité des ressources* en vérifiant si l'opération d'attribution des modes de communication est réversible (les stratégies AVA et VAV s'opposent l'une à l'autre). En référence à la notion d'*additivité*, c'est la *« commutativité »* de l'opération d'attribution des modes de communication qui était testée.

Les résultats observés ne sont pas allés dans ce sens. Dans l'expérience 4, l'une des stratégies bimodales (VAV) a donné lieu à un rappel plus faible. Cette stratégie a produit un effet négatif sur le rappel qui allait à l'encontre de l'hypothèse d'additivité des ressources. De plus, les évaluations subjectives de charge cognitive n'indiquaient pas d'augmentation de la charge dans cette condition. Ce résultat permet de conclure que *l'opération d'attribution des modes de communication n'est pas commutative*. L'effet de bi-modalité n'est pas validé.

Pour cette raison, l'expérience 4 indique qu'il est préférable de rejeter l'hypothèse d'additivité des ressources visuelles et auditives. Plus généralement, l'ensemble des résultats obtenus dans les expériences de la thèse indique que les sources de variation de la charge cognitive diffèrent des sources de variation de la performance d'apprentissage. Ces résultats permettent de conclure que le rôle causal de la notion de charge cognitive est faible au sein d'une tâche de DHM. A partir de ces résultats, et dans la mesure où les prédictions issues du point de vue pragmatique étaient plus précises, il est possible d'affirmer que les résultats contradictoires au sujet de l'*effet de modalité* (voir Ginns, 2005) et des autres effets identifiés dans le cadre de ces théories pourraient être réinterprétés à partir du point de vue pragmatique *pour en fournir l'explication*. Dans ce cas, comme en DHM, la prise en compte d'une diversité d'indicateurs (y compris comportementaux) sera nécessaire.

[1] Ces auteurs ajoutent une seconde voie à la partie gauche de cette chaîne, dans laquelle l'*expertise* à également une influence sur les ressources cognitives nécessaires, du fait qu'elle influence la *charge cognitive 'intrinsèque'*. Avec cet ajout, le lien entre *ressources cognitives* et *apprentissage* (partie droite) demeure identique (Gerjets & Scheiter, 2003, p. 35). Globalement, la chaîne causale est la même.

8.2.3 Conséquences théoriques

Ces résultats, ainsi que les implications qui en ont été tirées pour le domaine du DHM et pour le domaine de l'étude des activités d'apprentissage multimédia, permettent de faire quelques commentaires sur les notions théoriques utilisées dans ces deux domaines. Les points abordés portent d'abord sur le statut explicatif accordé à la notion de *charge cognitive*, puis sur la notion d'*information* telle qu'elle est utilisée actuellement en psychologie cognitive, puis enfin, quelques mots sont ajoutés sur l'intérêt de la notion d'*action*.

Ces remarques portent sur l'approche analytique en psychologie, dont l'objectif est de décomposer le processus étudié en ses composantes primaires (approche élémentariste) et les liens entre ces éléments (approche structuraliste) pour établir les relations entre ces éléments (approche causale). L'objectif est de discuter l'intérêt de ces notions dans une approche psychologique de la *communication homme machine*.

A La notion de charge cognitive

- ***Avant propos***

Le sujet d'origine de la thèse était :

> *« Constitution (en fonction d'indices) d'un modèle prédictif de la surcharge de travail mentale et de sa répartition en environnement multimodal. »*

Le modèle en question devait être appliqué au dialogue, contexte dans lequel le problème qui se pose est celui de la prise de décision pour la *'gestion de l'interaction'* et en vue du *« traitement des informations »*[1] par l'utilisateur (compréhension, mémorisation et réaction). De ce fait, le sujet a été reformulé autour d'une *problématique appliquée* de *conception des énoncés*, alors que le sujet d'origine relevait, au contraire, d'une *problématique théorique* de *modélisation de la charge*. Cependant, cette problématique appliquée n'a pas écarté la problématique théorique. Elle a, au contraire, permis de l'envisager sous un jour différent dans la mesure où la focalisation exclusive sur les processus cognitifs liée à la modélisation de la charge cognitive a été élargie à un ensemble des processus cognitifs et comportementaux qui pouvaient être pris en compte pour la conception des énoncés. Cette problématique de conception a permis de questionner les actions du système. Elle a impliqué d'étudier les effets des actions aussi largement que possible. En référence à la théorie de la collaboration (Clark, 1996), le souci était de prendre en compte un point de vue sociocognitif plutôt qu'un point de vue strictement cognitif.

Comme on l'a vu, cet élargissement de point de vue a permis d'obtenir des résultats qui remettent en question le rôle causal de la notion de charge cognitive. Cette idée avait été formulée ainsi après quelques mois de thèse : *« Il faut substituer à la notion de charge mentale/cognitive comme facteur explicatif central/général, une théorie analytique de l'action*

[1] Cette expression est utilisée ici dans la mesure où elle est consensuelle dans le contexte de la psychologie actuelle (*i.e.* en psychologie cognitive).

de l'individu en contexte. » Cette idée, bien que n'étant plus l'idée maîtresse de la thèse, en reste le fondement[1]. Elle est cohérente, notamment, avec le point de vue d'Austin (1962, *Cf.* chapitre 1) et son souci d'une théorie de l'action, ainsi qu'avec celui de Clark (1996) comme cela vient d'être noté.

- ***Résultat : la charge cognitive n'occupe pas une position causale centrale***

Les expériences de la thèse ont mis en évidence une diversité de causalités particulières liées au déroulement de l'activité et aux implications des actions, plutôt qu'une chaîne causale générale impliquant la charge cognitive.

Comme cela a été indiqué en conclusion de la synthèse des résultats (page 218), les effets des actes se déployaient sur plusieurs *niveaux hiérarchiques* et *niveaux de profondeur* des traitements (automatismes, modes de contrôle, réalisation de la tâche et performance cognitive). Ces résultats sont en cohérence avec les principes d'amorçage automatique du *modèle d'alignement interactif* (Pickering & Garrod, 2004) et renvoient, là aussi, à un modèle d'action. Ils remettent en cause l'intérêt central de la notion de charge cognitive et font écho aux critiques fréquentes de cette notion (de Montmollin, 1997 ; Meyer & Kieras, 1997 : Theureau, 2002 ; etc.), qui ont été évoquées dans le chapitre 4.

Pour poser la question de l'intérêt de la notion de charge, il est nécessaire de se reposer sur la distinction entre *analyse* et *synthèse*[2]. En effet, l'analyse est une opération de décomposition d'un tout en ses éléments, alors que la synthèse est l'opération inverse, qui consiste à décrire des éléments disjoints comme un tout. Il doit être clair que la notion de charge cognitive ne doit pas accéder au statut de *notion d'analyse*. Il n'existe aucun processus psychologique qui puisse être nommé *'charge cognitive'*. Il ne s'agit pas d'un élément qui pourrait être agrégé à d'autres pour constituer un tout. Il s'agit simplement d'une notion générale, qui peut permettre de qualifier un tout qui a été constitué par ailleurs. Il s'agit donc seulement, et uniquement, d'une *notion de synthèse*.

Dans ce cas, l'intérêt de cette notion est d'apporter des éléments généraux sur la tâche étudiée. C'est pourquoi les techniques de mesure utilisées (*Cf.* Tableau 4-1, p. 106) présentent de l'intérêt. Elles apportent des indications générales sur la tâche qui peuvent permettre de décider si une analyse est nécessaire ou non.

- ***Une « généralité vague »***

Ces indications générales peuvent être utiles car elles permettent de décider si la démarche de recherche des causes (l'analyse) est nécessaire. Mais l'utilisation de la notion de charge

[1] Un pas supplémentaire a été avancé dans la thèse puisque des éléments ont été fournis dans le sens de l'analyse des actions.

[2] La différence entre *analyse* et *synthèse* est présente chez de nombreux auteurs en philosophie, même anciens. On la trouve par exemple chez Descartes ou même chez Aristote. Mais la distinction formelle de ces deux notions n'est apparue qu'avec Kant (*Cf.* dictionnaire des concepts philosophiques : Blay, Castel, Engel, & Lenclud, 2006).

cognitive dans cette analyse tomberait sous le coup d'une accusation de sophisme[1], sous le nom de « généralité vague » proposé par Bentham (1824) :

> *« Les généralité vagues comprennent un large ensemble de sophismes que commettent ces personnes qui, aux expressions les plus particulières et les plus spécifiques autorisées par la nature du cas dont il s'agit, en préfèrent d'autres, plus générales et indéterminées. Une expression est vague et ambiguë, lorsqu'elle désigne par un seul et même nom, un objet qui peut être bon ou mauvais selon les circonstances. Si, au cours d'une enquête concernant les qualités d'un même objet, on utilise cette sorte d'expression, sans reconnaître ce qui les distingue, l'expression agit comme un sophisme »* (Bentham, 1824, p. 295)

Comme cela a été indiqué dans le chapitre 4, au cours de la présentation de la *théorie de la charge cognitive*, la formulation de cette théorie (Sweller, 1988) repose sur ce type de généralisation abusive. En effet, les commentaires apportés dans cette partie (*Cf.* pp. 113-114) ont permis de montrer que l'auteur avait d'abord proposé une analyse fonctionnelle et une formalisation des tâches qu'il étudiait. C'est dans un second temps qu'il a proposé une synthèse se basant sur la notion de charge cognitive. Dans cette synthèse, les fonctions dégagées dans l'analyse ont simplement fait l'objet d'un comptage. Tout ce que chacune d'entre elles avait de spécifique (*i.e.* leur utilité relative dans l'activité) était perdu dans ce comptage. Pour cette raison, on peut penser que la formulation de la théorie de la charge cognitive relève du sophisme de généralité vague. Différents processus ont été mis à jour dans l'analyse, avant d'être agglomérés.

- ***Illustration***

Pour illustrer cette remarque et tenter une réflexion d'ordre épistémologique, une *analogie* peut s'avérer utile. Le point de comparaison proposé est en provenance de la médecine et porte sur la lente évolution des connaissances qui ont accompagné la pratique de *la saignée*.

En effet, la saignée a été pratiquée de façon assez raisonnable (si l'on peut dire) jusqu'à la renaissance, puis elle a connue des développements extraordinaires qui l'ont conduite à son âge d'or, du XVIIe au XIXe siècle ; ceci alors même que les connaissances scientifiques étaient en plein développement (Beauchamp, 2000; Héritier, 1987). Notamment, la circulation du sang dans le corps a été décrite en 1628 (William Harvey), celle du système lymphatique en 1651 (Jean Pecquet), la physiologie des poumons en 1661 (Marcello Malpighi), leur rôle d'oxygénation du sang en 1669 (Richard Lower), l'existence des globules rouges en 1673 (Antoon van Leeuwenhoek), etc. Malgré ces découvertes, et contre des critiques fréquentes (*e.g.* chez Rabelais au XVIe siècle ; Maïmonide, dès le XIIe siècle, etc.) la saignée faisait encore l'objet de traités enflammés et d'une pratique intense au début du XIXe siècle (*e.g.* Broussais, 1816). La cause évoquée lors de la prescription consistait, notamment, à qualifier le patient de *'pléthorique'* (*i.e.* en surabondance de sang), ce qui imposait de faire baisser la

[1] Bentham (1824, p. 183) définit ainsi la notion de sophisme : *« On désigne ordinairement du nom de "sophisme" tout argument avancé ou tout sujet de discussion suggéré afin de, ou avec la probabilité de produire l'effet de tromper ou de causer quelque opinion erronée susceptible d'être admise par toute personne dont l'esprit a pu se trouver mis en présence de cet argument. »*

pression, selon un schéma quantitatif qui peut, aujourd'hui, prêter à rire, mais qui a pourtant fait de nombreux morts, avec beaucoup de sérieux… La saignée était prescrite pour la quasi-totalité des maladies, y compris un simple rhume, et même pour les bien-portants, à titre préventif (*Cf.* Beauchamp, 2000; Héritier, 1987). Elle était le plus souvent prescrite de bonne foi, pour des raisons qui tiennent seulement au niveau d'avancement des connaissances scientifiques propre à l'époque. Elle était prescrite malgré des recensements de cas et l'application de métriques destinées à en mesurer l'efficacité, dès le XVIe siècle, qui plaidaient contre son indication (Héritier, 1987) ; et malgré les descriptions anatomiques et les découvertes qui, sans l'interdire, ne plaidaient pas en sa faveur. Ce n'est finalement qu'en pénétrant dans l'*infiniment petit* (biochimie du sang) que les préjugés à l'origine de la pratique ont pu être définitivement levés. Cela ne date que du début du XXe siècle avec l'apparition de l'*hématologie* et la mise en évidence des *groupes sanguins*.

L'analogie est assez directe. La volonté de réduire la charge cognitive des utilisateurs renvoie à l'idée d'une cognition *'pléthorique'*. Le manque de connaissances concernant les processus en cause (la *physiologie du sang* dans le cas de la saignée, la *physiologie du sens* dans le cas de la charge cognitive) permet d'envisager le phénomène selon une conception quantitative, voire linéaire. L'objectif pour la recherche scientifique est de dépasser cette première approximation. Pour cela, il est nécessaire de décomposer le phénomène et d'y faire toutes les distinctions nécessaires pour produire une analyse des processus impliqués. C'est le cas du *modèle de Wickens* (*Cf.* Figure 4-2, p. 98), qui permet de distinguer les pools de ressources cognitives les plus importants. Mais cette distinction n'est pas suffisante. La distinction des types d'information, utilisée dans les expériences 3, 4 et 5, a permis de proposer une catégorisation supplémentaire. Cela a permis de mettre en évidence le rôle essentiel de la *'relation type-mode'* dans l'action, conformément aux remarques de Clark & Brennan (1991) qui soulignaient l'importance de cette relation. En effet, cette notion permet de dépasser celle de *'code'* et d'envisager le fonctionnement cognitif comme un processus plus riche qu'un simple décodage d'informations. Les interprétations d'un observateur sur les actes d'une personne (ou même d'un objet) sont beaucoup plus riches que l'*information* qui est supposée être traitée dans la théorie cognitive classique. La *relation type-mode* permet de qualifier les actes selon plusieurs dimensions, ce qui permet d'envisager cette richesse et de dépasser une vision quantitative unidimensionnelle du fonctionnement cognitif, sous l'angle de la charge. Il en ressort que les formalismes cognitifs peuvent être considérablement enrichis. Mais en attendant un point d'aboutissement de ces développements, il est toujours interdit d'interdire la notion de charge cognitive [1]. Cela viendra peut-être avec le développement des techniques de l'intelligence artificielle, quand elles permettront de modéliser de façon plus uniforme et homogène l'*infiniment petit cognitif*[2].

[1] Et puis nous, psychologues, nous ne tuons personne…

[2] Cette expression peut être trompeuse. Elle ne fait pas référence spécifiquement au fonctionnement du cerveau, mais aux mécanismes attributifs mis en place par l'individu (pris dans sa globalité). Le fonctionnement du cerveau est inclus dans cette globalité. D'autres aspects, tels que la hiérarchisation

- ***Une métaphore alternative***

Barrouillet (1996) évoque deux métaphores qui permettent d'envisager différemment les ressources cognitives : (1) la *métaphore spatio-temporelle* et (2) la *métaphore énergétique* (*Cf.* page 93). Dans le premier cas, la partie active de la mémoire de travail est vue comme l'espace occupé dans l'espace disponible. Dans le second cas, la partie active est vue comme l'ensemble des connaissances qui ont atteint un certain seuil d'activation.

Les résultats de l'expérience 4 ont permis d'introduire une notion alternative qui permet de proposer une troisième métaphore. La notion utilisée était celle de *« métaphore vibratoire »* (ou *« métaphore ondulatoire »*). En effet, le cerveau peut être décrit par des phénomènes d'excitations ondulatoires (on parle d'ondes cérébrales). Il a pu être décrit comme un ensemble de modules spécialisés en communication (Fodor, 1983). La théorie classique indique que le résultat du traitement d'un module est utilisé en entrée du module suivant (voir Richard, 2004), ce qui suppose un fonctionnement en série, analysable selon une approche élémentariste. Or, la structure du cerveau est très complexe et implique des connexions multiples qui agissent en parallèle (*e.g.* Gil, 2006). L'analyse de ce système repose alors sur le principe de Pascal qui indique qu'il n'est pas interdit *« pour connaître le tout de connaître également les parties, non plus que pour connaître les parties, de connaître également le tout. »* Selon ce principe, il est nécessaire de considérer l'organisation de la structure étudiée. L'ensemble des *pools de ressources* disponibles sont actifs en continu au cours de l'activité et ils exercent une influence mutuelle les uns sur les autres en fonction des variations des paramètres de la tâche. Les réactions, automatiques et contrôlées, mises en place par l'individu dépendraient alors des influences mutuelles d'un ensemble de variations externes, liées à la tâche, et d'un ensemble de variation internes, liées à l'individu. Dans ce système, chaque acte est susceptible de produire des effets divers, qui doivent être estimés spécifiquement pour prédire la performance.

Dans cette perspective, le fonctionnement du système dépend de la *résonance*[1] des mécanismes d'activation et d'inhibition en provenance de chacun des modules. Ce fonctionnement renvoie à l'équilibration des mécanismes dans le système, conformément au principe d'homéostasie. Le problème ne serait pas, alors, d'estimer la quantité de ressources actives dans la tâche, mais d'estimer si les ressources activées chez un individu sont susceptibles d'atteindre un état de résonance suffisant pour permettre à cet individu de répondre aux exigences de la tâche dans des conditions acceptables.

Il peut être noté que l'ouvrage de référence de Norbert Wiener (*"Cybernetics"*, 1948) faisait appel à un tel principe de *résonance*. Mais celui-ci a été oublié dans le cadre de la psychologie cognitive, notamment parce que cette discipline a été basée sur la notion d'*information*, conformément aux propos d'un autre auteur (Shannon, 1948).

---

temporelle de l'interaction (externe à l'individu) ainsi que les connaissances conventionnelles (interindividuelles) y sont également inclus.

[1] *Résonance* (définition du Robert) : *« Augmentation de l'amplitude d'un système physique en vibration lorsque la vibration excitatrice se rapproche d'une fréquence naturelle de ce système. »*

B La notion d'information

Le chapitre 4 a permis d'examiner les prémisses de la notion de capacité limitée (pp. 94-96) et de montrer que dans la *théorie de l'information*, tout comme dans la *théorie de la mémoire de travail*, l'hypothèse de base consiste à assimiler toute information dans une catégorie unique *« information »*. Les arguments de Miller (1956) sont très explicites dans ce sens (*Cf.* p. 96). Ils sont présentés dès le début de l'article et développés dans tout le corps du texte. Pour cette seule raison, les arguments tirés de la *différence analyse-synthèse* et du *sophisme de généralité vague* pourraient être appliqués à l'emploi de la notion d'information en psychologie. Ces points ne sont pas développés puisqu'ils l'ont été au sujet de la notion de charge cognitive.

La notion de spécificité modale mise en évidence grâce aux expériences 3 et 4 a permis de montrer qu'il existe une *relation entre le type d'information et le mode de présentation.* Cette relation stipule que les informations n'ont pas toutes les mêmes propriétés. Trois types d'information ont été différenciés (*'écho'*, *'réponse'*, *'relance'*). Ils ont été associés en se basant sur les deux fonctions principales en dialogue (*'contrôle'* et *'information'*) pour former deux groupes : le groupe des *« informations de contrôle »* et le groupes des *« informations d'information »*[1]. Les informations issues de ces deux groupes ont provoqué des effets différents dans la communication :

- Les *informations de contrôle* étaient moins bien détectées quand elles étaient présentées sur la modalité visuelle. La modalité auditive a permis une réactivité plus importante ;
- Les *informations d'information* étaient négociées trop rapidement et donnaient lieu à un contrôle permanent qui nuisait à la consultation. La modalité visuelle a permis une consultation plus sereine, qui a permis une meilleure mémorisation.

Ces résultats indiquent que des effets différents ont été obtenus non seulement en termes de mémorisation, mais également en termes d'interactions (deux plans conventionnels distincts dans la Figure 5-1). Il existerait donc au moins deux types d'actes différents ; ce qui permet de conclure que l'utilisation du terme *« information »* pour faire référence à une catégorie indifférenciée est abusive.

L'article de Baber et al. (1996) précisait :

> *« Considérées globalement les études de la théorie des ressources multiples suggèrent qu'un modèle viable du traitement humain de l'information peut être développé dans lequel on considère que le traitement de l'information dépend des codes utilisés pour présenter l'information. »* (p. 40, traduction libre)

Les expériences 3, 4 et 5 ont permis de montrer que la relation entre le code utilisé et les effets qui en résultent n'est pas aussi directe. Il existe des actes de différents types qui

[1] Cette répétition est un indicateur de l'ambiguïté du terme *information*. Dans la perspective pragmatique, il serait préférable de parler d'*acte de contrôle* et d'*acte d'information*. La répétition a été conservée ici pour souligner cette ambiguïté.

engendrent des effets différents bien qu'ils relèvent du même code. Cette conclusion s'applique au minimum pour les actes verbaux, les seuls utilisés dans les expériences de la thèse. – D'autres travaux expérimentaux seraient utiles pour la généralisation. – Dans la définition de Nigay et Coutaz (1993), la notion de *« modalité »* associe un langage d'interaction (*i.e.* un code) et un dispositif technique. Ainsi, dans l'expression *relation type-mode* (ici, « mode » et « modalité » sont considérés comme synonymes), le terme *« mode »* renvoie, entre autres, au « code » utilisé pour présenter l'information. Mais l'expression complète intègre également le terme « type » qui renvoie quand à lui à une catégorisation des informations qui ne dépend pas du code. Les résultats obtenus permettent d'affirmer contre Baber et al. (1996) qu'un modèle du traitement humain de l'information qui se baserait sur les codes utilisés pour présenter l'information serait insuffisant et, par suite, qu'un tel modèle n'est pas viable. La notion de *code* est insuffisante pour qualifier les actes.

Le *traitement humain de l'information* est un peu plus complexe. Il dépend des effets que le producteur du message souhaite engendrer et de la *'résonance'* de ses actes dans l'espace fonctionnel du destinataire (son système cognitif). Cet espace fonctionnel est lié aux différentes *boucles perception-action* qu'est capable de mettre en œuvre ce destinataire, en fonction de ses capacités physiques et mentales, soit en fonction de son comportement et de ses connaissances. Pour cette raison, la notion d'information est trompeuse car elle laisse supposer qu'il existerait une permanence du contenu du message. En réalité, le contenu est toujours relatif à l'emploi qui est fait (*« en ces circonstances, à cet auditoire, dans ce dessein et cette intention. »*, Austin, 1962).

- ***Pour une « Théorie de la Relativité Epistémique »***

En ce sens, il doit être entendu que la signification d'un message (*i.e.* son *épistémè*. *Cf.* Foucault, 1966)[1] est relative aux finalités de l'action en cours et à tous les éléments de la situation susceptibles d'être utilisés pour l'interprétation des actions.

La majeure partie de ces éléments a été représentée dans le schéma de synthèse proposé au début de la partie expérimentale (Figure 5-1, p. 124). Comme cela a été expliqué dans la partie consacrée aux stratégies de présentation en DHM (dans ce chapitre, *Cf.* pp. 219-223), l'interprétation d'un acte repose (au moins) sur la prise en compte simultanée des trois *plans d'intentions* (*communicatives*, *assertives* et *interactives*) qui ont été identifiés. Dans ce contexte, l'*information* fournie par un message n'est pas « contenue » dans l'énoncé correspondant. La signification est construite par l'interprète (*Cf.* Merleau-Ponty, 1945) à partir de l'énoncé, qui ne joue qu'un rôle de *marqueur locutoire* (Wilson, 1997). L'*énoncé* n'est qu'un acte. Il reste indéterminé en l'absence d'un interprète. Ainsi, l'expression *« traitement des informations »* est un non-sens puisqu'il ne préexiste aucune information à traiter…

[1] La notion d'*épistémè* renvoie à l'ensemble des connaissances propres à un groupe social et à une époque. L'emploi qui est fait de ce terme ici restreint le groupe à l'équipe en communication et l'époque au moment de la communication. Dans ce cas, l'*épistémè* renvoie à l'ensemble des connaissances relatives à ce groupe et ce moment.

Selon ce point de vue, la signification de l'énoncé dépend des intentions du locuteur et de leur interprétation par le destinataire. D'abord, quand un locuteur produit un énoncé grâce auquel il exprime une certaine proposition, la signification de sa proposition dépend de ses intentions. Par exemple, la signification d'une phrase telle que : *« On te soutiendra à mort »*[1] dépend des intentions du locuteur d'apporter effectivement son soutien ou non (*intentions communicatives* et *interactives*, en particulier). Par ailleurs, l'auditeur qui perçoit l'énoncé interprète les intentions du locuteur. Il a la possibilité de sélectionner ou de privilégier certaines intentions. Par exemple, un énoncé aussi banal que : *« Passe-moi le sel »* peut avoir des significations différentes selon que l'interprète privilégie l'interprétation communicative (*e.g. « Il faut toujours que tu donnes des ordres ! »* ou *« Et c'est quoi le mot magique ? »*), l'interprétation assertive (*e.g. « Le sel de Guérande ou le Baleine ? »*) ou l'intention interactive (*e.g. « Regarde. Il est sous ton nez. »*). De même, les énoncés d'un système de DHM font l'objet d'interprétations sur ces différents plans (*e.g.* respectivement : (1) il parle dans le vide, (2) il répète des choses que je sais et (3) il est lent).

A fortiori, la signification dépendra également du contexte, puisque les intentions du locuteur et les interprétations du destinataire dépendent elles-mêmes du contexte. Par exemple, la phrase *« Je suis le roi du monde »* (*« I am the King of the world »*) n'a pas le même sens si elle est prononcée par un acteur jouant une scène d'un film, par le réalisateur du film lors d'une cérémonie de remise de prix, ou par un touriste qui reproduirait la scène en question au cours d'une traversée *Cherbourg-Cork*[2]. Dans ces trois cas, la signification de l'énoncé est différente car le même contenu propositionnel entretien des relations différentes avec les *intentions communicatives*, *assertives* et *interactives*. Globalement, ces exemples permettent de rappeler que la signification d'un acte est relative aux buts reconnus par l'interprète[3].

Ainsi, la distinction de plusieurs plans d'intentions permet de montrer que la signification dépend au minimum des *intentions communicatives, assertives* et *interactives* qui sont interprétées. L'*information* peut être déduite à partir des constructions interprétatives. Elle correspond au différentiel entre l'état des connaissances de l'interprète avant et après l'interprétation. Ainsi, la Figure 5-1, permet de proposer une *'théorie de la relativité épistémique'* qui stipule la relativité du sens aux buts qui motivent l'action, aux buts pris en compte pour l'interpréter et aux finalités contextuelles.

Dans ce cas, la notion d'information ne peut être utilisée comme une unité d'analyse car elle n'est pas première. L'information est le résultat de l'interprétation. Elle est relative à un individu et, au sens strict, elle ne peut pas être objectivée. Toute information supposée objective est nécessairement issue de la confrontation de plusieurs points de vue subjectifs.

[1] Voir : http://www.dailymotion.com/video/x4c9x4

[2] Le lecteur qui ne comprendrait pas cet exemple est invité à visionner le film *Titanic* (de James Cameron et avec Leonardo Di Caprio).

[3] L'interprète peut être le *destinataire*, le *participant* ou l'*auditeur*, mais aussi le *locuteur* lui-même. Par exemple, quand un locuteur produit un *lapsus* il peut y avoir une divergence entre ses intentions initiales et l'interprétation qu'il fait de ses actes.

Les causes de l'interprétation, qui permettraient d'expliquer le phénomène/processus attributif, doivent être recherchées dans d'autres notions que celle d'information.

C La notion d'action

La notion d'*action* peut être substituée à celle d'*information* pour fournir l'analyse attendue. Austin (1962) n'est pas le seul auteur qui ait plaidé en faveur de cette notion. Elle a été théorisée par d'autres auteurs importants, dont notamment Léontiev (1977) et Vygotsky (1978). Des auteurs récents ont rappelé le rôle essentiel de l'action dans le fonctionnement psychologique (*e.g.* Glenberg, 1997; Varela, 1988). Par ailleurs, des travaux de réhabilitation de ces approches ont été proposés récemment dans une série d'articles (Bedny & Harris, 2008; Bedny & Karwowski, 2003; Bedny, Karwowski & Bedny, 2001). Ces auteurs ont examiné les liens entre théorie de l'activité (d'origine russe) et théorie de l'action (d'origine allemande) et proposent d'appliquer ces approches dans le contexte de la conception des systèmes d'interaction homme machine. Dans ces approches, le but est d'expliquer le comportement de l'individu dans le contexte de son activité en prenant en compte tous les éléments nécessaires.

Ces propositions tirent également partie des apports du point de vue constructiviste (Piaget, 1967, 1970), qui a permis de reformuler le paradigme psychologique. Dans ce sens, Piaget indiquait, par exemple :

> *« L'instrument d'échange initial n'est pas la perception, comme les rationalistes l'ont trop facilement concédé à l'empirisme, mais bien l'action elle-même en sa plasticité beaucoup plus grande. Certes, les perceptions jouent un rôle essentiel, mais elles dépendent en partie de l'action en son ensemble et certains mécanismes que l'on aurait pu croire innés ou très primitifs (…) ne se constituent qu'à un certain niveau de la construction des objets. De façon générale, toute perception aboutit à conférer aux éléments perçus des significations relatives à l'action (…) et c'est donc de l'action qu'il convient de partir. »* (Piaget, 1970, p. 12)

Cette citation appuie fortement les remarques proposées ci-dessus au sujet d'une *« théorie de la relativité épistémique »*. Plus généralement, les épistémologies constructivistes sont diverses et permettent toujours de spécifier la relativité du sens à l'action (Le Moigne, 2007). Les travaux proposés dans la thèse se veulent en accord avec ces approches car elles fournissent, comme cela a été largement commenté dans les pages qui précèdent, des outils d'analyse utiles et puissants, qui permettent une description fine de l'activité, sur plusieurs niveaux, et à l'aide desquels les conceptions théoriques peuvent être éprouvées.

8.2.4 Pragmatique et Pertinence

La *théorie des actes de langage* (Austin, 1962) est fondamentale car elle fournit l'analyse de base à partir de laquelle l'approche pragmatique a pu se développer. Comme on l'a vu, cette approche a été la plus précise pour prédire les résultats des expériences de la thèse.

Les résultats obtenus valident l'intérêt de cette approche pour étudier la *présentation multimodale des informations* et pour *interpréter l'ensemble des processus* qui opèrent dans l'activité. Les ressources individuelles d'un utilisateur, y compris cognitives, ont des propriétés pour l'action. L'approche pragmatique permet d'aborder ces propriétés sous un angle descriptif, sans *a priori* théorique quant à la finalité générale du fonctionnement cognitif (*e.g.* mémoire, développement, apprentissage). Elle permet d'envisager les processus cognitifs dans leur dimension de *traitement symbolique abstrait des actes*. De ce point de vue, la Figure 5-1 tente de représenter (et d'organiser) les connaissances qu'un individu doit mettre en relation pour faire ce traitement symbolique. Dans ce sens, ce schéma pourrait être utilisé pour inspirer des études et des modélisations computationnelles exploitant les finesses des théories et modèles pragmatiques.

Dans cette perspective, les conclusions proposées ici sont en accord avec le point de vue défendu par différents auteurs (*e.g.* Johnstone, Berry, Nguyen, & Asper, 1994 ; Schegloff, 2006) qui indiquent que le *DHM* n'est pas fondamentalement différent du *dialogue entre humains* et que les interactions occupent une place fondamentale. Le problème qui se pose alors pour la recherche scientifique est celui de l'étude et de l'explication de la richesse comportementale dont sont capables des humains. Il s'agit d'expliquer la compétence pragmatique pour tenter d'en reproduire le fonctionnement dans des systèmes artificiels.

- ***La pertinence***

La notion de pertinence est au cœur de cette problématique. Le chapitre 1 a permis de rappeler que le *principe de pertinence* et le rapport entre *effet* et *effort* proposé par Sperber et Wilson (1989) ont été rejeté par différents auteurs (*e.g.* Allwood, 1984 ; 1995 ; Caron, 1989 ; Schegloff, 2006) qui renvoient aux conventions conversationnelles et à l'aspect relationnel de la pertinence, qui est à la fois *multi-niveaux* et *multifonctionnelle*.

Il s'agit d'une notion générale à l'aide de laquelle il est possible de qualifier l'utilité relative des actes du locuteur dans la communication et leur cohésion avec les obligations conversationnelles, qui sont multiples (Allwood, 1995). La pertinence définit l'utilité positive des actes, au service de la communication. Elle renvoie donc aux finalités de ce processus dans le contexte de la communication (*i.e.* effets contextuels, voir Sperber & Wilson, 1989).

Cette notion est souvent considérée comme importante pour le développement des systèmes de dialogue. Par exemple, Landragin (2003) s'est appuyé sur les propositions de Sperber et Wilson (1989) pour tenter de formaliser la pertinence dans le cadre du fonctionnement d'un système de DHM multimodal. Mais la problématique du développement technique des systèmes renvoie d'abord à des processus de niveau inférieur (liés à la perception) qui ont conduit cet auteur à évoluer vers la notion de saillance, plus proche des facteurs physiques (voir, Landragin, 2004). Cette tentative d'application indique que les notions d'*effet* et d'*effort* conduisent à des difficultés opérationnelles dissuasives.

Dans le même esprit que Sperber et Wilson (1989), une formulation *en intension* de la pertinence peut être proposée. En effet, les notions d'*effet* et d'*effort* étaient proposées en

référence aux informations supposées être véhiculées par l'énoncé (informations communicative et informative, *Cf.* chapitre 1). Sur ce principe, le rapport *'effet/effort'* renvoyait à une position ontologique qui se voulait objective (*i.e.* positiviste, ou encore réaliste). Or, comme cela a été indiqué, la notion d'action renvoie, au contraire, à une position constructiviste qui est plus cohérente avec le paradigme psychologique moderne, issu notamment des apports de Piaget (1967). L'action se rapporte aux buts de l'individu qui agit. Dans ce cas, des buts peuvent être associés aux différents niveaux de coordination proposés par Allwood (1995) et Clark (2002). Il est alors possible de proposer le rapport suivant :

Pertinence = Action / Fonction

Mais une telle écriture ne doit pas être prise au premier degré. La pertinence n'est pas une valeur numérique et cette écriture ne renvoie pas à un rapport mathématique. Encore une fois, la pertinence est un concept relationnel (Allwood, 1995). *Actions* et *fonctions* sont des éléments discrets et leur formalisation dans des systèmes artificiels pose avant tout le problème de la reconnaissance des actes (capacité de catégorisation) et de l'identification de leur utilité (capacité de raisonnement causal finaliste). Ainsi, cette formule doit être dépliée sur les différents niveaux de coordination et sur les différents plans d'intentions pour évaluer si l'action réalisée répond aux fonctions (*i.e.* aux buts) qu'elle vise.

La pertinence renvoie à une capacité à produire de la signification en contexte. Plus encore, dans la communication entre humains, elle renvoie à la capacité à générer des énoncés informatifs, *i.e.* des énoncés qui amènent les auditeurs à produire des inférences nouvelles pour eux. Pour cette raison, la formalisation de la pertinence dans des systèmes artificiels suppose de donner à ces systèmes la possibilité de construire une représentation des inférences construites par l'interlocuteur (modèle de l'utilisateur), pour y répondre au mieux.

8.3 Limites et perspectives

La plupart des remarques qui peuvent être faites sur les limites et perspectives de la thèse ont été faites au fil de la discussion. Elles sont brièvement rappelées dans cette partie.

8.3.1 Limites

- ***Magicien d'Oz***

L'utilisation du protocole en *Magicien d'Oz* a permis de poser des hypothèses sur le DHM dans des services très formatés. Dans ces expériences, la diversité des capacités d'action du système restait relativement limitée. Cette limitation présentait l'avantage de permettre des comparaisons plus faciles entre les réactions des participants, ce qui est utile dans un contexte expérimental. Mais le risque associé à ce type de restriction est que les effets observés soient restreints au protocole utilisé. Un travail de généralisation est donc

nécessaire pour appliquer les résultats obtenus dans un cadre de la conception de systèmes de DHM réels, dans le contexte de services commerciaux destinés au grand public.

L'implémentation de systèmes en fonctionnement utilisant les principes d'analyse qui ont été utilisés pour concevoir les expériences de la thèse permettrait d'éprouver ces principes et de les intégrer parmi les contraintes prises en compte au cours du processus de développement. Ce type de travail serait l'occasion d'une collaboration interdisciplinaire qui permettrait de traiter de façon systématique toutes les questions liées à l'implémentation.

- ***Utilisation du code verbal***

Les expériences proposées n'ont porté que sur le code verbal. Comme cela a été indiqué, l'existence d'une forme auditive de ce code (la parole) et d'une forme visuelle (l'écriture) a permis de proposer des stratégies de présentation audio-visuelles opposables entre elles, ce qui a permis de mettre en avant la notion de *spécificité modale*. Mais cela a aussi limité le matériel expérimental à ce code.

Pour cette raison, la notion de *« type d'acte »* devrait être éprouvée dans d'autres contextes codiques (ainsi que dans d'autres contextes d'activité) pour évoluer vers une catégorisation aussi universelle que possible.

- ***Terminologie***

Ce point n'a pas été évoqué auparavant.

Pour l'analyse psychologique, le nombre de dimensions à prendre en compte est important. La métaphore d'un espace tridimensionnel (ou même tétra-dimensionnel) est insuffisante. Notamment, les trois dimensions d'Austin (locutoire, illocutoire et perlocutoire) sont des macro-dimensions sous lesquelles peuvent être rangées tous les types de connaissances de l'individu (connaissances sur soi, connaissances sur les conventions, connaissances sur les individus et le monde). Ces connaissances peuvent être scindées en sous-dimensions (interactif, assertif, social), lesquelles peuvent être représentées chacune par des espaces complexes.

Les imbrications de systèmes de dimensions sont nombreuses et il serait utile de disposer d'un vocabulaire permettant de décrire différents type de relations d'imbrication.

8.3.2 Perspectives

- ***Dialogue Homme Machine***

Des indications ont été apportées pour l'étude des énoncés du système en DHM. Le développement de cette problématique renvoie au problème de la formalisation des connaissances du système. Ce point ne concerne pas seulement la représentation de l'utilisateur. Il renvoie au problème plus général des ontologies, *i.e.* de l'organisation des connaissances du système en général (représentation de la langue, de la tâche, des utilisateurs, d'un dialogue etc.)

Ces aspects renvoient à des problèmes psychologiques, mais ne peuvent être étudiés dans leur dimension dynamique (fondamentale) sans les possibilités techniques permises par le développement de l'intelligence artificielle. En effet, la complexité du système impose de tester les jeux de relations grâce à ces techniques. Cela laisse entrevoir la possibilité de travaux interdisciplinaires riches dans les années à venir. Dahlbäck (2003) renvoie cette question à un problème de formation des chercheurs, qui doivent être capables de prendre en considération simultanément des contraintes issues de plusieurs traditions disciplinaires.

- ***Effet de modalité et apprentissage***

Comme cela a été indiqué dans la partie consacrée à l'étude des énoncés en situation d'apprentissage (pp. 224-225), les résultats autour de l'*effet de modalité* et des *effets des 'formats de présentation'* en général pourraient être réinterprétés dans une théorie de l'action pour tenter d'expliquer les résultats contradictoires souvent notés par les auteurs (voir Gerjets & Scheiter, 2003 ; Ginns, 2005).

- ***Pour une psychologie pragmatique***

Du point de vue psychologique, les résultats obtenus et la synthèse proposée (Figure 5-1) donnent lieu à des réflexions sur des notions et distinctions préexistantes telles que, comme cela a déjà été noté, les notions de *charge cognitive* et d'*information*. Un point peut être ajouté sur la distinction entre connaissances *procédurales* et *déclaratives* (Anderson, 1983).

Les auteurs qui utilisent cette distinction dans leurs travaux tendent à ranger les informations dans l'une ou l'autre de ces deux catégories (exclusivement de l'autre) pour faire des hypothèses sur leurs effets. Mais ce mode de classement est arbitraire et il amène à des difficultés lors de la construction du matériel expérimental (voir à ce sujet, Babin, 2007). Comme on l'a vu, Goffman (1987) insistait sur la dimension *mot-geste* de la parole. En effet, l'énoncé est un acte qui a une dimension déclarative (*intentions assertives*) ainsi qu'une dimension procédurale (*intentions interactives*). Ces deux dimensions coexistent, entre elles et avec d'autres (*Cf.* Figure 5-1), relativement à tout acte. Il est important de considérer que les effets des actes ne sont pas exclusifs mais agissent en parallèle. La *psychologie cognitive*, qui analyse le fonctionnement psychologique en se basant sur la notion d'information n'est pas apte à rendre compte de ces aspects. Pour cela, une psychologie se basant sur la notion d'action est nécessaire. Dans ce cas, il serait intéressant de faire émerger plus explicitement une *psychologie pragmatique*. Cela peut être le cas, par exemple, dans le cadre du développement de la *psychologie ergonomique* qui, comme cela a été indiqué, se veut être *« une psychologie de la performance »*.

Cette distinction déclaratif/procédural est, notamment, à la base de l'architecture du modèle ACT-R (Anderson et al., 2004). Les remarques faites ici indiquent que des approches plus complètes sont envisageables. Ainsi, un équivalent européen à ACT-R serait le bienvenu. En effet, dans le contexte international chaque continent a tout intérêt à prendre des décisions d'action qui servent les intérêts qui lui sont propres (*i.e.* qui sont *'pertinentes'* pour lui). Par ailleurs, en dehors de ce genre de considération, il serait intéressant pour la communauté

psychologique dans son ensemble de disposer de solutions différentes, susceptibles d'envisager des principes de fonctionnement et d'utiliser des techniques de modélisation qui ne seraient pas focalisées uniquement sur les approches américaines les plus consensuelles.

A partir du schéma de la Figure 5-1, des propositions peuvent être faites pour évoluer vers des systèmes capables, par exemple, de simuler certains aspects de la personnalité, ou de construire des représentations riches sur l'utilisateur, ces aspects se déployant nécessairement dans l'action. De même, cette figure peut être utilisée pour interpréter des cas de quiproquo (*e.g.* si différents interprètes ordonnent différemment les différents plans d'intentions). Ces éléments peuvent être intéressants dans le cadre du développement d'une *robotique génétique* (*e.g.* Oudeyer, Kaplan & Hafner, 2007) car ils peuvent permettent de structurer et de faire fonctionner un système de connaissances, au service de l'action.

« En science et surtout en politique, les idées, souvent plus têtues que les faits, résistent au déferlement des données et des preuves. »

Edgar Morin

« J'ai découvert combien il est vain de ne polémiquer que contre l'erreur : celle-ci renaît sans cesse de principes de pensée qui, eux, se trouvent hors conscience polémique. J'ai compris combien il était vain de prouver seulement au niveau du phénomène : son message est bientôt résorbé par des mécanismes d'oubli qui relèvent de l'auto-défense du système d'idées menacé. J'ai compris qu'il était sans espoir de seulement réfuter : seule une nouvelle fondation peut ruiner l'ancienne. C'est pourquoi je pense que le problème crucial est celui du principe organisateur de la connaissance, et ce qui est vital aujourd'hui, ce n'est pas seulement d'apprendre, pas seulement de réapprendre, pas seulement de désapprendre, mais de réorganiser notre système mental pour réapprendre à apprendre*. »*

Edgar Morin

« Ce qui apprend à apprendre, c'est cela la méthode. »

Edgar Morin

« La solitude à laquelle je me suis contraint est le lot du pionnier, mais aussi de l'égaré. J'ai perdu le contact avec ceux qui n'ont pas entrepris le même voyage et je ne vois pas encore mes compagnons qui existent, sans doute, et qui eux non plus ne me voient pas… »

Edgar Morin

BIBLIOGRAPHIE

Allemandou, J. (2007). *SIMDIAL : Un paradigme d'évaluation automatique de systèmes de dialogue homme-machine par simulation déterministe d'utilisateurs.* Unpublished Thèse, LIMSI/CNRS, Paris.

Allen, J. (1983). Maintaining Knowledge about Temporal Intervals. *Communications of the ACM, 26*(11), 832-843.

Allen, J. (1995). *Natural Language Processing* (2nd ed.). Redwood, CA: Benjamin Cummings Publishing Company.

Allport, D. A. (1980). Attention and Performance. In G. L. Claxton (Ed.), *Cognitive psychology: New directions* (pp. 112-153). London: Routledge & Kegan Paul.

Allwood, J. (1976). *Linguistic Communication as Action and Cooperation.* Unpublished Gothenburg Monographs in Linguistics 2, University of Göteborg, Göteborg.

Allwood, J. (1977). A Critical Look at Speech Act Theory. In Dahl (Ed.), *Logic, Pragmatics and Grammar* (pp. 53-69). Lund: Studentlitteratur.

Allwood, J. (1984). On Relevance in Spoken Interaction. In Bäckman & Kjellmer (Eds.), *Papers on Language and Literature* (pp. 18-35): Acta Universitatis Gothoburgensis.

Allwood, J. (1995). An Activity Based Approach to Pragmatics. In H. Bunt & B. Black (Eds.), *Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics* (Vol. 76, pp. 47-80). Amsterdam: John Benjamins.

Allwood, J. (2006). Treebanks for Spoken Language - Some Reflections. In P. J. Henrichsen & P. R. Skadhauge (Eds.), *Treebanking for Discourse and Speech* (Vol. Copenhagen Studies in Language, 32, pp. 29-42). Copenhagen: Samfundslitteratur Press.

Allwood, J. (2007). Activity Based Studies of Linguistic Interaction. In *Gothenburg Papers in Theoretical Linguistics*: Göteborg University, Dept. of Linguistics.

Allwood, J., Nivre, J., & Ahlsén, E. (1992). On the Semantics and Pragmatics of Linguistic Feedback. *Journal of Semantics, 9*(1), 1-26.

Allwood, J., Traum, D., & Jokinen, K. (2000). Cooperation, dialogue and ethics. *Human-Computer Studies, 53*, 871-914.

Amalberti, R., Carbonnell, N., & Falzon, P. (1993). User representations of computer systems in human-computer speech interaction. *Man-Machine Studies, 38*, 547-566.

Amiel, V. (2005). *Peut-on parler de Collaboration en Dialogue Homme-Machine ?* Unpublished Thèse, Université Toulouse II, Toulouse.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist, 51 (4)*, 355-365.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111 (4)*, 1036-1060.

Anderson, J. R., & Bower, G. H. (1973). *Human Associative memory*. Washington: Winston and Sons.

Anderson, J. R., Qin, Y., Sohn, M.-H., Stenger, V. A., & Carter, C. S. (2003). An information-processing model of the BOLD response in symbol manipulation tasks. *Psychonomic Bulletin & Review, 10 (2)*, 241-261.

Andre, A. D. (2001). The Value of Workload in the Design and Evaluation of Consumer Products. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload and fatigue: theory, research and practice* (pp. 373-383). New Jersey: Lawrence Erlbaum.

Austin, J. L. (1962). *How to do Things with Words*: Oxford University Press (Version française : (1970) *Quand dire c'est faire.* Edition du Seuil. Paris).

Baber, C., & Mellor, B. (2001). Using critical path analysis to model multimodal human-computer interaction. *International Journal of Human-Computer Studies, 54*, 613-636.

Baber, C., Mellor, B., Graham, R., Noyes, J. M., & Tunley, C. (1996). Workload and the use of automatic speech recognition: The effects of time and resource demands. *Speech Communication, 20*, 37-53.

Babin, L.-M. (2007). *Aides à l'optimisation de l'apprentissage d'un système interactif.* Unpublished Thèse, Université Toulouse II Le Mirail, Toulouse.

Bachvarova, Y., van Dijk, B., & Nijholt, A. (2007). Towards a Unified Knowledge-Based Approach to Modality Choice. In I. v. d. Sluis, M. Theune, E. Reiter & E. Krahmer (Eds.), *Proceedings Workshop on Multimodal Output Generation (MOG 2007)* (pp. 5-15). Aberdeen, Scotland.

Baddeley, A. D. (2004). The Psychology of Memory. In A. D. Baddeley, M. D. Kopelman & B. A. Wilson (Eds.), *The Essential Handbook of Memory Disorders for Clinicians.* (pp. 1-13): John Wiley & Sons, Ltd.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-90). San Diego: Academic Press.

Baddeley, A. D., & Logie, R. H. (2003). The Multiple-Component Model. In A. Miyake & P. Shah (Eds.), *Models of Working Memory - Mechanisms of Active Maintenance and Executive Control* (pp. 28-61). Cambridge: Cambridge University Press.

Baker, M. (2003). Les dialogues avec, autour et au travers des technologies éducatives. *L'Orientation Scolaire et Professionnelle, 32*(3), 359-397.

Baker, M. (2004). *Recherches sur l'élaboration de connaissances dans le dialogue.* Université de Nancy, Nancy.

Balota, D. A., Engle, R. W., & Cowan, N. (1990). Suffix Interference in the Recall of Linguistically Coherent Speech. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16 (3)*, 446-456.

Bangerter, A., Clark, H. H., & Katz, A. R. (2004). Navigating Joint Projects in Telephone Conversations. *Discourse Processes, 37*(1), 1-23.

Bard, E. G., Anderson, A. H., Chen, Y., Nicholson, H. B. M., Havard, C., & Dalzel-Job, S. (2007). Let's you do that: Sharing the cognitive burdens of dialogue. *Journal of Memory and Language, 57*, 616–641.

Bateson, G. (1973). *Vers une Ecologie de l'Esprit (titre original : Steps to an Ecology of Mind)*. Paris: Editions du Seuil.

Baudoin, F. (2007). *Personnalisation des Systèmes de Dialogue en Langage Naturel : une Méthode d'Anticipation Rationnelle d'Actions Communicatives.* Unpublished Thèse, Université Paris 6 - Lip6, Paris.

Beauchamp, C. (2000). *Le sang et l'imaginaire médical - Histoire de la saignée au XVIIIe et XIXe siècles* (Esculape ed.). Paris: Desclée de Brouwer.

Bedny, G., & Harris, S. R. (2008). "Working sphere/engagement" and the concept of task in activity theory. *Interacting with Computers, 20*(2), 251-255.

Bedny, G., & Karwowski, W. (2003). A Systemic-Structural Activity Approach to the Design of Human–Computer Interaction Tasks. *International Journal of Human-Computer Interaction, 16*(2), 235–260.

Bedny, G., Karwowski, W., & Bedny, M. (2001). The Principle of Unity of Cognition and Behavior: Implications of Activity Theory for the Study of Human Work. *International Journal of Cognitive Ergonomics, 5*(4), 401–420.

Bentham, J. (1824). *Manuel de sophismes politiques*. Paris: Bruylant L.G.D.J.

Bernsen, N. O. (1994). Foundations of multimodal representations. A taxonomy of representational modalities. *Interacting with Computers, 6*, 347-371.

Bernsen, N. O., Dybkjaer, H., & Dybkjaer, L. (1998). *Designing Interactive Speech Systems: From First Ideas to User Testing.* New-York, NY: Springer Verlag.

Berthold, A., & Jameson, A. (1999). *Interpreting Symptoms of Cognitive Load in Speech Input.* Paper presented at the User Modeling: Proceedings of the Seventh International Conference, UM99., Vienna, New York: Springer Wien New York.

Black, W., Allwood, J., Bunt, H., Dols, F., Donzella, C., Ferrari, G., et al. (1991). A Pragmatics Based Language Understanding System. In *Proceedings of the ESPRIT Conference*.

Blanchet, P. (1995). *La pragmatique, d'Austin à Goffman* (Bertrand Lacoste ed.). Paris.

Blay, M., Castel, P.-H., Engel, P., & Lenclud, G. (Eds.). (2006). Paris: Larousse.

Bratman, M., Israel, D., & Pollack, M. (1988). Plans and resource-bounded practical reasoning. *Computational Intelligence, 4*(2), 349-355.

Brennan, S. E., & Clark, H. H. (1996). Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22 (6)*(6), 1482-1493.

Brennan, S. E., & Hulteen, E. (1995). Interaction and feedback in a spoken language system: A theoretical framework. *Knowledge-Based Systems, 8*, 143-151.

Bretier, P. (1995). *La communication orale coopérative : contribution à la modélisation logique et à la mise en oeuvre d'un agent rationnel dialoguant.* Unpublished thèse, Université Paris Nord, Paris.

Bretier, P., Le Bigot, L., Panaget, F., & Sadek, D. (2004). *De la représentation de l'interlocuteur vers un modèle utilisateur formel pour le dialogue personne-machine.* Paper presented at the Proceedings of IHM'2004, International Conference Proceedings Series, Namur (Belgium).

Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence, 47*, 139-159.

Broussais, F. (1816). *Examen de la doctrine médicale généralement adoptée et des systèmes modernes de nosologie*. Paris: Maronval.

Bubb-Lewis, C., & Scerbo, M. W. (2002). The effects of communication modes on performance and discourse organization with an adaptive interface. *Applied Ergonomics, 33*, 15-26.

Buisine, S., & Martin, J.-C. (2007). The effects of speech-gesture cooperation in animated agents' behavior in multimedia presentations. . *Interacting with Computers, 19*, 484-493.

Bunt, H. (1994). Context and Dialogue Control. *THINK Quarterly*.

Bunt, H. (2000). Dialogue pragmatics and context specification. In H. Bunt & W. Black (Eds.), *Abduction, Belief and Context in Dialogue. Studies in Computational Pragmatics.* (pp. 81-150). Amsterdam: John Benjamins.

Bunt, H., Kipp, M., Maybury, M., & Wahlster, W. (2003). Fusion and Coordination for Multimodal Interactive Information Presentation. In O. Stock & M. Zancanaro (Eds.), *Multimodal Intelligent Information Presentation* (pp. 325-339). Dordrecht Springer.

Byrne, M. D., & Bovair, S. (1997). A working memory model of a common procedural error. *Cognitive Science, 21*, 31-61.

Cahn, J. E., & Brennan, S. E. (1999). A psychological model of grounding and repair in dialog. In *Proceedings, AAAI Fall Symposium on Psychological Models of Communication in*

Collaborative Systems (pp. 25-33). North Falmouth, MA: American Association for Artificial Intelligence.

Camus, J.-F. (1998). L'attention. In J. L. Roulin (Ed.), *Psychologie Cognitive* (pp. 138-205). Paris: Bréal.

Card, S., Moran, T. P., & Newell, A. (1983). *The Psychology of Human Computer Interaction.*: Lawrence Erlbaum Associates.

Caron, J. (1989). *Précis de psycholinguistique*. Paris: Presses Universitaires de France.

Case, R. (1992). *The mind's staircase: Exploring the conceptual underpinnings of children's thought and knowledge*. Hillsdale, NJ: Lawrence Erlbaum.

Cassell, J., Kopp, S., Tepper, P., Ferriman, K., & Striegnitz, K. (2007). Trading Spaces: How Humans and Humanoids use Speech and Gesture to Give Directions. In T. Nishida (Ed.), *Conversational Informatics* (pp. 133-160). New York: John Wiley & Sons.

Castelfranchi, C. (2007). For a systematic theory of expectations. In S. Vosniadou, D. Kayser & A. Protopapas (Eds.), *Proceedings of EuroCogSci07, the European Cognitive Science Conference* (pp. 10-17). Delphes: Lawrence Erlbaum Associates.

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction, 8(4)*, 293-332.

Chandler, P., & Sweller, J. (1996). Cognitive Load While Learning to Use a Computer Program. *Applied Cognitive Psychology, 10*, 151-170.

Chanquoy, L., Tricot, A., & Sweller, J. (2007). *La Charge Cognitive*.

Chapanis, A., Ochsman, R. B., Parrish, R. N., & Weeks, G. D. (1972). Studies in Interactive Communication: I. The Effects of Four Communication Modes on the Behavior of Teams During Cooperative Problem-Solving. *Human Factors, 14 (6)*, 487-509.

Chapanis, A., & Overbey, C. M. (1974). Studies in interactive communication: III. Effects of similar and dissimilar communication channels and two interchange options on team problem solving. *Perceptual and Motor Skills, 38*, 343-374.

Chapanis, A., Parrish, R. N., Ochsman, R. B., & Weeks, G. D. (1977). Studies in Interactive Communication: II. The Effects of Four Communication Modes on the Lignuistic Performance of Teams During Cooperative Problem Solving. *Human Factors, 19 (2)*, 101-126.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Clark, H. H. (2002). Speaking in time. *Speech Communication, 36*, 5-13.

Clark, H. H. (2004). Pragmatics of Language Performance. In R. Horn & G. Ward (Eds.), *Handbook of pragmatics* (pp. 365-382). Oxford: Blackwell.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. Resnick, J. M. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC: APA.

Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language, 58*, 332-373.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*, 62-81.

Clark, H. H., & Marshall, C. (1978). Reference diaries. In D. L. Waltz (Ed.), *Theoretical issues in natural language processes, TINLAP-2* (pp. 57-63). New-York: Association for Computing Machinery.

Clark, H. H., & Marshall, C. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. Webber & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1989). Contributing To Discourse. *Cognitive Science, 13*, 259-294.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1-39.

Clarke, T., Ayres, P., & Sweller, J. (2005). The Impact of Sequencing and Prior Knowledge on Learning Mathematics Through Spreadsheet Applications. . *Educational Technology Research & Development, 53*(3), 15-24.

Clémente, P. (2004). *Vers la formalisation des capacités multimodales d'un agent rationnel dialoguant.* Université de Franche-Comté, Besançon.

Cohen, P. R., & Levesque, H. J. (1990). Intention is Choice with Commitment. *Artificial Intelligence, 42*, 213-261.

Coutaz, J., & Nigay, L. (1994). Les propriétés "CARE" dans les interfaces multimodales. In *Actes de la conférence IHM'94* (pp. 7-14). Lille.

Crowder, R. G. (1967). Prefix effects in immediate memory. *Canadian Journal of Psychology, 21*, 450-461.

Dahlbäck, N. (2003). If cognitive science is multidisciplinary, which are the disciplines? Cognitive science as three methodological cultures. In F. Schmalhofer (Ed.), *Proceedings of EuroCogSci'03* (pp. 5). Osnabrück, Germany: Lawrence Erlbaum Associates.

Dahlbäck, N., Jonsson, A., & Ahrenberg, L. (1993). Wizard of OZ studies - why and how. In *Proceedings Intelligent User Interfaces'93* (pp. 193-200).

Damnati, G., Béchet, F., & De Mori, R. (2007). Experiments on the France Telecom 3000 Voice Agency corpus: academic research on an industrial spoken dialog system. In *Bridging the Gap: Academic and Industrial Research in Dialog Technologies Workshop Proceedings April 2007* (pp. 48-55). NAACL-HLT, Rochester, NY: Association for Computational Linguistics.

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9*, 561-584.

De Montmollin, M. (1997). *Vocabulaire de l'ergonomie*. Toulouse: Editions Octarès.

Deng, L., & Huang, X. (2004). Challenges in adopting speech recognition *Communications of the ACM, 47 (1)*(Special issue: Multimodal interfaces that flex, adapt, and persist), 69-75.

Devooght, K. (2007). Modélisation de la capacité en fonction des activités d'un agent intentionnel. In *Actes des journées d'Intelligence Artificielle Fondamentale 2007 (IAF'07)*. Grenoble: Actes électroniques (http://www.cril.univ-artois.fr/~konieczny/IAF07/).

Diao, Y., & Sweller, J. (2007). Redundancy in foreign language reading comprehension instruction: Concurrent written and spoken presentations. *Learning and Instruction, 17*, 78-88.

Duclaye, F. (2003). *Apprentissage automatique de relations d'équivalence sémantique à partir du Web.*, ENST - INFRES, Brest.

Dumazeau, C. (2005). *Favoriser l'établissement d'un contexte mutuellement partagé dans les communications distances.* Unpublished Thèse, Centre d'étude de la navigation aérienne (CENA), Toulouse.

Edlund, J., Gustafson, J., Heldner, M., & Hjalmarsson, A. (2008). Towards human-like spoken dialogue systems. *Speech Communication, Accepted 06/04/2008*, 24.

Engle, R. W. (1974). The modality effect: Is precategorical acoustic storage responsible? *Journal of Experimental Psychology, 102*, 824-829.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypothesis. *Journal of Experimental Psychology: Learning, Memory and Cognition, 5*, 972-992.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102 (2)*, 211-245.

Fitts, P. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*, 381-391.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge: MIT Press.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language*: McGraw Hill.

Foucault, M. (1966). *Les mots et les choses. Une archéologie des sciences humaines.* Paris: Gallimard.

Foucault, M. (1976). *Histoire de la sexualité I : La volonté de savoir* (tel ed.). Paris: Gallimard.

Fraser, N. M., & Gilbert, G. N. (1991). Simulating speech systems. *Computer Speech and Language, 5*, 81-99.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 8*(1), 8-11.

Gazdar, G. (1979). *Pragmatics: Implicature, Presupposition and Logical Form*. New-York: Academic Press.

Gerjets, P., & Scheiter, K. (2003). Goal Configurations and Processing Strategies as Moderators Between Instructional Design an Cognitive Load: Evidence From Hypertext-Based Instruction. *Educational Psychologist, 38 (1)*, 33-41.

Gibbon, D., Moore, R., & Winski, R. (1997). *Handbook of Standards and Resources for Spoken Language Systems.* New-York, NY: Mouton de Gruyter.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Gil, R. (2006). *Neuropsychologie* (4ème édition ed.). Paris: Masson.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*, 313-331.

Glass, J., & Weinstein, E. (2001). SpeechBuilder: Facilitating Spoken Dialogue System Development. In *Proceedings of the 7th European Conference on Speech Communication and Technology*. Aalborg Denmark.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20*, 1-55.

Goffman, E. (1974). *Les rites d'interaction*. Paris: Les Editions de Minuit.

Goffman, E. (1987). *Façons de parler*. Paris: Les Editions de Minuit.

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: why bother with subjective measures? *Human Factors, 26*, 519-532.

Gordon, D., & Lakoff, G. (1975). Conversational Postulates. In C. e. Morgan (Ed.), *Syntax and Semantics (V. 3)* (pp. 83-106).

Gorin, A. L., Parker, B. A., Sachs, R. M., & Wilpon, J. G. (1996). How May I Help You? In *Proceedings on Interactive Voice Technology for Telecommunications Applications (IVTTA)* (pp. 57-60).

Grice, P. (1957). Meaning. *Philosophical Review, 66*, 377-388.

Grice, P. (1969). Utterer's meaning and intentions. *Philosophical Review, 78*, 147-177.

Grice, P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics: Speech Acts* (Vol. 3, pp. 41-58). New-York: Academic Press.

Guhe, M. (2007). Towards a cognitive model of multimodal output for language production. In I. v. d. Sluis, M. Theune, E. Reiter & E. Krahmer (Eds.), *Proceedings of Workshop on Multimodal Output Generation (MOG 2007)* (pp. 59–68). Aberdeen, Scotland.

Hapeshi, K., & Jones, D. (1992). Interactive Multimedia for Instruction: A Cognitive Analysis of the Role of Audition and Vision. *International Journal of Human-Computer Interaction, 4 (1)*, 79-99.

Haviland, S. E., & Clark, H. H. (1974). Psychological processes as llnguistic explanation. In D. Cohen (Ed.), *Phenomena*. Washington: Hemisphere Publishing Corp.

Haviland, S. E., & Clark, H. H. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension.* Norwood, NJ: Ablex.

Helleberg, J. R., & Wickens, C. D. (2003). Effects of Data-Link Modality and Display Redundancy on Pilot Performance: An Attentional Perspective. *The International Journal of Aviation Psychology, 13(3)*, 189-210.

Héritier, J. (1987). *La sève de l'homme*. Paris: Denoël.

Hernault, H., Piwek, P., Prendinger, H., & Ishizuka, M. (2008). Generating Dialogues for Virtual Agents Using Nested Textual Coherence Relations. In *Proceedings of IVA08: 8th International Conference on Intelligent Virtual Agents*. Tokyo, Japan.

Hitch, G. J. (1975). The role of attention in visual and auditory suffix effects. *Memory & Cognition, 3*, 501-505.

Hoc, J.-M., & Amalberti, R. (2007). Cognitive Control dynamics for reaching a satisficing performance in complex dynamic situations. *Manuscript submitted for publication. Journal of Cognitive Engineering and Decision Making, 1*(1), 22-55.

Hoc, J.-M., & Darses, F. (2007). *Projet de création de groupement de recherche « Psycho Ergo » : Psychologie ergonomique et Ergonomie cognitive*: CNRS (sections 27, 34, 07).

Hone, K. S., & Baber, C. (1995). Using a simulation method to predict the transaction time effects of applying alternative levels of constraint to utterances within speech interactive dialogues. In , . In P. Dalsgaard, L. Larsen, L. Boves & I. Thomsen (Eds.), *Proceedings of the ESCA Workshop on Spoken Dialogue Systems* (pp. 209-212). Vigso, Denmark.

Horchani, M. (2007). *Vers une Communication Humain-Machine Naturelle : Stratégies de Dialogue et de Présentation Multimodales.* Unpublished Thèse, Université Joseph Fourier - Grenoble 1, Grenoble.

Horchani, M., Fréard, D., Caron, B., Jamet, E., Nigay, L., & Panaget, F. (2007). *Stratégie de dialogue et de présentation multimodale : un composant logiciel dédié et son application à des expérimentations en Magicien d´Oz.* Paper presented at the conférence IHM´07, Paris (France).

Horchani, M., Fréard, D., Jamet, E., Nigay, L., & Panaget, F. (2007a). *Applying experimental results on cognitive load to a multimodal dialogic strategy component.* Paper presented at the Interact´07.

Horchani, M., Fréard, D., Jamet, E., Nigay, L., & Panaget, F. (2007b). A Platform for Designing Multimodal Dialogic and Presentation Strategies. In *The 2007 Workshop on the Semantics and Pragmatics of Dialogue (DECALOG)* (pp. 169-170).

Horchani, M., Nigay, L., & Panaget, F. (2007). A Platform for Output Dialogic Strategies in Natural Multimodal Dialogue Systems. In *Proc. of the IUI* (pp. 206-215). Honolulu, Hawaii.

Horton, W. S. (in press). The influence of partner-specific memory associations on language production: Evidence from picture naming. *Language and Cognitive Processes, 23/24*, xxx-xxx.

Horton, W. S., & Gerrig, R. J. (2005a). The impact of memory demands on audience design during language production. *Cognition*, 16.

Horton, W. S., & Gerrig, R. J. (2005b). Conversational Common Ground and Memory Processes in Language Production. *Discourse Processes*, 67.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*, 91-117.

Horton, W. S., & Spieler, D. H. (2007). Age-Related Differences in Communication and Audience Design. *Psychology and Aging, 22*(2), 281-290.

Howell, M., Love, S., & Turner, M. (2006). Visualisation improves the usability of voice-operated mobile phone services. *International Journal of Human-Computer Studies, 64*, 754-769.

Huguenard, B. R., Lerch, F. J., Junker, B. W., Patz, R. J., & Kass, R. E. (1997). Working-Memory Failure in Phone-Based Interaction. *ACM Transactions on Computer-Human Interaction, 4*, 67-102.

Iqbal, S. T., Zheng, X. S., & Bailey, B. P. (2004). *Task-evoked pupillary response to mental workload in human-computer interaction.* Paper presented at the Conference on Human Factors in Computing Systems archive, Vienna, Austria.

Jackendoff, R. (2007). Linguistics in Cognitive Science : The state of the art. *Linguistic review, 24*(4), 347-401.

Jacko, J. A., Moloney, K. P., Kongnakorn, T., Barnard, L., Edwards, P. J., Leonard, V. K., et al. (2005). Multimodal Feedback as a Solution to Ocular Disease-Based User Performance Decrements in the Absence of Functional Visual Loss. *International Journal of Human-Computer Interaction, 18*(2), 183 - 218.

Jameson, A., Kiefer, J., Müller, C., Großmann-Hutter, B., Wittig, F., & Rummer, R. (2006). Assessment of a user's time pressure and cognitive load on the basis of features of speech. *Journal of Computer Science and Technology*.

Jamet, E. (2004). Apprentissage et multimédias pédagogiques : quelques méthodes en psychologie expérimentale. In E. de Vries (Ed.), *Comment évalue-t-on les technologies informatiques pour l'apprentissage ?* (pp. 54-70).

Jamet, E., & Le Bohec, O. (2007). The effect of redundant text in multimedia instruction. *Contemporary Educational Psychology, 32*(4), 588-598.

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science, 4*, 71-115.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge, UK: Cambridge University Press.

Johnstone, A., Berry, U., NGuyen, T., & Asper, A. (1994). There was a long pause: influencing turn-taking behaviour in human-human and human-computer spoken dialogues. *International Journal of Human-Computer Studies, 41*, 363-411.

Jokinen, K., & Hurtig, T. (2006). User Expectations and Real Experience on a Multimodal Interactive System. In *Proceedings of Interspeech 2006*. Pittsburgh, US.

Jokinen, K., & Raike, A. (2003). Multimodality – technology, visions and demands for the future. In *The 1st Nordic Symposium on Multimodal Interfaces* (pp. 12). Copenhagen.

Juang, B. H., & Rabiner, L. R. (2004). Automatic Speech Recognition: A Brief History of the Technology [Electronic Version]. *www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/354_LALI-ASRHistory-final-10-8.pdf*, 24.

Just, M. A., & Carpenter, P. A. (1992). A Capacity Theory of Comprehension : Individual Differences in Working Memory. *Psychological Review, 99 (1)*, 122-149.

Just, M. A., Carpenter, P. A., & Miyake, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric, and event-related potential studies of brain work. , , . . In *Theoretical Issues in Ergonomics Science* (Vol. 4, pp. 56-88). Special Edition.

Juul-Kristensen, B., Laursen, B., Pilegaard, M., & Jensen, B. R. (2004). Physical workload during use of speech recognition and traditional computer inpu devices. *Ergonomics, 47 (2)*, 119-133.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs: NJ: Prentice Hall.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. *Educational Psychologist, 38 (1)*, 23-31.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40*, 1-17.

Karsenty, L. (2000). MECI : Une méthode d'explicitation des modèles utilisateurs implicites. In *Actes de la conférence ERGO-IHM 2000* (pp. 8). Biarritz, 3-6 Octobre.

Karsenty, L. (2003). *Ergonomie cognitive des communications : la question du contexte partagé.* Unpublished Habilitation à Diriger des Recherches, Université René Descartes - Paris V, Paris.

Keizer, S., & Bunt, H. (2006). Multidimensional dialogue management. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue* (pp. 37-45). Sydney: ACM.

Keizer, S., & Bunt, H. (2007a). Evaluating combinations of dialogue acts. In *Proceedings of the Eighth SIGDIAL Workshop on Discourse and Dialogue (SIGDIAL 2007)* (pp. 158-165). Antwerp.

Keizer, S., & Bunt, H. (2007b). Evaluating combinations of dialogue acts for generation. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue* (pp. 158-165). Antwerp, Belgium.

Kennedy, A., Wilkes, A., Elder, L., & Murray, W. S. (1988). Dialogue with machines. *Cognition, 30(1)*, 37-72.

Kerbrat-Orecchioni, C. (1995). *Les interactions verbales*. Paris: Armand Colin Editeur.

Kester, L., Lehnen, C., Van Gerven, P. W. M., & Kirschner, P. A. (2006). Just-in-time, schematic supportive information presentation during cognitive skill acquisition. *Computers in Human Behavior, 22*(1), 93-112.

Keysar, B., & Paek, T. S. (1993). Definite reference and mutual ignorance. In *The 34th annual meeting of the Psychonomics Society*. Washington, DC.

Kieras, D. E., & Meyer, D. E. (1997). An Overview of the EPIC Architecture for Cognition and Performance with Application to Human-Computer Interaction. *Human-Computer Interaction, 12 (4)*, 391-438.

Kieras, D. E., & Meyer, D. E. (2000). The Role of Cognitive Task Analysis in the Application of Predictive Models of Human Performance. In J. M. Shraagen, S. F. Chipman & V. L. Shalin (Eds.), *Cognitive Task Analysis* (pp. 237-260). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Kieras, D. E., Meyer, D. E., Mueller, S., & Seymour, T. (2003). Insights into Working Memory from the Perspective of the EPIC Architecture for Modeling Skilled Perceptual-Motor and Cognitive Human Performance. In A. Miyake & P. Shah (Eds.), *Models of Working Memory - Mechanisms of Active Maintenance and Executive Control* (pp. 183-223). Cambridge: Cambridge University Press.

Krauss, R. M., & Fussell, S. R. (1989). Other-relatedness in language processing: Discussion and comments. *Journal of Language and Social Psychology, 7*, 263-279.

Krauss, R. M., & Glucksberg, S. (1969). The development of communication: Competence as a function of age. *Child Development, 40*, 255-256.

Krauss, R. M., & Glucksberg, S. (1977). Social and non-social speech. *Scientific American, 236*, 100-105.

Krauss, R. M., & Pardo, J. S. (2004). Is alignment always the result of automatic priming ? *Behavioral & Brain Sciences, 27*, 203-204.

Lahlou, S. (2002). Travail de bureau et débordement cognitif. In M. Jourdan & J. Theureau (Eds.), *Charge mentale : notion floue et vrai problème* (pp. 73-91). Toulouse: Octarès.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago.

Landragin, F. (2003). *Modélisation de la Communication Multimodale - Vers une Formalisation de la Pertinence.* Unpublished Thèse, Université Poincaré, Nancy.

Landragin, F. (2004). Saillance physique et saillance cognitive. . *Cognition, Représentation, Langage (CORELA), 2*(2), http://edel.univ-poitiers.fr/corela/document.php?id=142.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.

Le Bigot, L., Jamet, E., & Rouet, J.-F. (2004). Searching information with a natural language dialogue system: A comparison of spoken vs. written modalities. *Applied Ergonomics, 35(6)*, 557-564.

Le Bigot, L., Jamet, E., Rouet, J.-F., & Amiel, V. (2006). Mode and modal transfer effects on performance and discourse organization with an information retrieval dialogue system in natural language. *Computers in Human Behavior, 22(3)*, 467-500.

Le Bigot, L., Rouet, J.-F., & Jamet, E. (2007). Effects of Speech- and Text-Based Interaction Modes in Natural Language Human-Computer Dialogue. *Human Factors, 49*(6), 1045-1053.

Le Bigot, L., Terrier, P., Amiel, V., Poulain, G., Jamet, E., & Rouet, J.-F. (2007). Effect of modality on collaboration with a dialogue system. *International Journal of Human-Computer Studies, 65*, 983-991.

Le Bodic, L. (2005). *Approche de l'évaluation des systèmes interactifs multimodaux par simulation comportementale située.* Unpublished thèse, Université de Bretagne Occidentale, Brest.

Le Bohec, O., & Jamet, E. (2005). Les effets de redondance dans l'apprentissage à partir de documents multimédia. *Le Travail Humain, 68 (2)*, 97-124.

Le Guilcher, B., & Villame, T. (2002). Conception des Interactions Individus.Systèmes : réflexions pour dépasser le critère de charge mentale. In M. Jourdan & J. Theureau (Eds.), *Charge mentale : Notion floue et vrai problème.* (pp. 93-119). Toulouse: Octarès.

Le Moigne, J.-L. (2007). *Les épistémologies constructivistes* (Que sais-je ? ed.). Paris: PUF.

Lefebvre. (2008). *Les indicateurs non verbaux dans les interactions médiatisées.* Université de Bretagne-Sud, Vannes.

Leiser, R. G. (1989). Exploiting convergence to improve natural language understanding. *Interacting with Computers, 1*(3), 284 - 298.

Léontiev, A. N. (1977). *Activity, consciousness and personality*. Moscou: Political Publishers.

Leplat, J. (2001). Eléments pour une histoire de la notion de charge mentale. In M. Jourdan & J. Theureau (Eds.), *Charge mentale : Notion floue et vrai problème* (pp. 27-40). Toulouse: Ocatrès.

Leplat, J. (2002). Eléments pour une histoire de la notion de charge mentale. In M. Jourdan & J. Theureau (Eds.), *Charge mentale : Notion floue et vrai problème* (pp. 27-40). Toulouse: Ocatrès.

Leplat, J., & Welford, A. T. (1978). Symposium on Mental Workload. *Ergonomics, Special issue*, 141-233.

Levelt, W. J. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-75.

Levinson, S. E., Rabiner, L. R., & Sondhi, M. M. (1983). An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition. *Bell Systems Technical Journal, 62*(4), 1035-1074.

Lewis, C., & Norman, D. A. (1986). Designing for Error. In D. A. Norman & S. W. Draper (Eds.), *User centered system design: New perspectives on Human-Computer Interaction* (pp. 411-432). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Lippmann, R. P. (1997). Speech recognition by machines and humans. *Speech Communication, 22*, 1-15.

Louis, V. (2002). *Conception et mise en oeuvre de modèles formels du calcul de plans d'action complexes par un agent rationnel dialoguant.* Université de Caen, Caen.

Lovett, M. C., Reder, L. M., & Lebiere, C. (2003). Modelling Working Memory in a Unified Architecture - An ACT-R perspective. In A. Miyake & P. Shah (Eds.), *Models of Working Memory - Mechanisms of Active Maintenance and Executive Control* (pp. 135-182). Cambridge: Cambridge University Press.

Markus, L. M. (1987). Toward a "Critical Mass" Theory of Interactive Media: Universal Access, Interdependence and Diffusion. *Communication Research, 14*(5), 491-511.

Martin, J.-C., Niewiadomski, R., Devillers, L., Buisine, S., & Pelachaud, C. (2006). Multimodal complex emotions: Gesture expressivity and blended facial expressions. *International Journal of Humanoid Robotics, Special Edition "Achieving Human-Like Qualities in Interactive Virtual and Physical Humanoids"*.

Matarazzo, J. D., Weitman, M., Saslow, G., & Wiens, A. N. (1963). Interviewer influence on duration of interviewee speech. *Journal of Verbal Learning and Verbal Behavior, 1*, 451-458.

Maury, J. P. (1991). *Recommandations aux partenaires des services vocaux téléphoniuqes*. Paris: France Télécom.

Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 31-48). Cambridge: Cambridge University Press.

Mayer, R. E., & Coll. (2005). *The Cambridge Handbook of Multimedia Learning*. Cambridge: Cambridge University Press.

Mayer, R. E., & Moreno, R. (1998). A Split-Attention Effect in Multimedia Learning: Evidence for Dual Processing Systems in Working Memory. *Journal of Educational Psychology, 90 (2)*, 312-320.

Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist, 38 (1)*, 43-52.

Mc Corquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical construct and intervening variables. *Psychological Review, 55*, 99-107.

Mc Tear, M. F. (2002). Spoken Dialogue Technology: Enabling the Conversational User Interface. *Computing Surveys, 34 (1)*, 90-169.

Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Paris: Gallimard.

Meyer, D. E., & Kieras, D. E. (1997). A Computational Theory of Executive Cognitive Processes and Multiple-Task Performance: Part 1. Basic Mechanisms. *Psychological Review, 104 (1)*, 3-65.

Meyer, G. (2006). *Formalisation logique de préférences qualitatives pour la sélection de la réaction d'un agent rationnel dialoguant.*, Thèse de l'Université de Paris XI, Paris.

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review, 3*, 81-97.

Minker, W., & Néel, F. (2002). Développement des technologies vocales. *Le Travail Humain, 65 (3)*, 261-288.

Moray, N. (1979). *Mental Workload, Theory and Measurement*. New York: Plenum.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358-368.

Moreno, R., & Mayer, R. E. (2002). Verbal Redundancy in Multimedia Learning: When Reading Helps Listening. *Journal of Educational Psychology, 94 (1)*, 156-163.

Moreno, R., & Mayer, R. E. (2007). Interactive Multimodal Learning Environments. *Educational Psychology Review, 19*, 309-326.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing Cognitive Load by Mixing Auditory and Visual Presentation Modes. *Journal of Educational Psychology, 87 (2)*, 319-334.

Müller, C., Großmann-Hutter, B., Jameson, A., Rummer, R., & Wittig, F. (2001). Recognizing Time Pressure and Cognitive Load on the Basis of Speech: An Experimental Study. In J. Vassileva, P. Gmytrasiewicz & M. Bauer (Eds.), *Proceedings of the Eighth International Conference UM2001 (User Modeling)*. Berlin.

Murray, A. C., Jones, D. M., & Frankish, C. R. (1996). Dialogue design in speech-mediated data-entry : the role of syntactic constraints and feedback. *International Journal of Human-Computer Studies, 45*, 263-286.

Navon, D. (1984). Resources-A theoretical soupstone? *Psychological Review, 91*, 216-234.

Navon, D., & Gopher, D. (1979). On the economy of the human information processing system. *Psychological Review, 86*, 214-255.

Nievergelt, J., & Weydert, J. (1980). Sites, Modes, and Trails: Telling the User of an interactive System Where he is, What he can do, and How to get places. In R. A. Guedj (Ed.), *Methodology of Interaction* (pp. 327-338). Amsterdam.

Nigay, L., & Coutaz, J. (1993). A Design Space for Multimodal Systems: Concurrent Processing and Data Fusion. In S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel & T. White (Eds.), *Proceedings of the ACM CHI 93 Human Factors in Computing Systems Conference* (pp. 172-178). Amsterdam, The Netherlands.

Norman, D. A. (1999). Affordances, conventions and design. *Interactions, 6 (3)*(3), 38-42.

Norman, D. A., & Bobrow, D. (1975). On data-limited and resource-limited processing. *Journal of Cognitive Psychology, 7*, 44-60.

Ochs, M., Niewiadomski, R., Pelachaud, C., & Sadek, D. (2006). Expressions Intelligentes des Emotions. *Revue en Intelligence Artificielle RIA, Special Edition "Interaction Emotionnelle", 20*(4-5).

Ombredane, A., & Faverge, J. M. (1955). *L'analyse du travail*. Paris: PUF.

Oudeyer, P.-Y., Kaplan, F., & Hafner, V. V. (2007). Intrinsic Motivation Systems for Autonomous Mental Development. *IEEE Transactions on Evolutionary Computation, 11*(2), 265-286.

Oviatt, S., Coulston, R., & Lunsford, R. (2004). *When Do We Interact Multimodally? Cognitive Load and Multimodal Communication Patterns.* Paper presented at the ICMI'04, State College, Pennsylvania, USA.

Oviatt, S., Darves, C., & Coulston, R. (2004). Toward adaptive conversational interfaces: Modeling speech convergence with animated personas. *ACM Transactions on Computer-Human Interaction, 11*(3), 300-328.

Paas, F. G. W. C., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist, 38 (1)*, 1-4.

Paivio, A. (1986). *Mental representations: A dual coding approach* (New York and Oxford Oxfordshire ed.). University of Western Ontario: Oxford University Press.

Panaget, F. (1996). *D'un système générique de génération d'énoncés en contexte de dialogue oral à la formalisation logique des capacités linguistiques d'un agent rationnel dialoguant.* Université de Rennes 1, Rennes.

Parasuraman, R., & Hancock, P. A. (2001). Adaptative Control of Mental Workload. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload and fatigue: theory, research and practice* (pp. 306-320). New Jersey: Lawrence Erlbaum.

Parush, A. (2005). Speech-Based Interaction in Multitask Conditions: Impact of Prompt Modality. *Human Factors, 47 (3)*, 591-597.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition, 17 (4)*, 398-422.

Piaget, J. (1967). *Logique Et Connaissance Scientifique* (Encyclopédie de la Pléiade ed.). Paris: Gallimard.

Piaget, J. (1970). *L'épistémologie génétique*. Paris: PUF.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*.

Pontal, J. F. (1997). *Recommandations pour la conception des services téléphoniuqes à commande vocale*. Paris: France Télécom.

Posner, M. I., & Boies, S. (1971). Componants of Attention. *Psychological Review, 78*, 391-408.

Posner, M. I., Nissen, J. M., & Klein, R. (1976). Visual dominance: An information processign account of its origins and significance. *Psychological Review, 83*, 157-171.

Potjer, J., Russel, A., Boves, L., & Os, E. D. (1996). Subjective and objective evaluation of two types of dialogues in a call assistance service. In *IVTTA, IEEE* (pp. 121-124). Basking Ridge, NJ.

Putnam, H. (1983). *Methodology, Epistemology, and Philosophy of Science: Essays in Honour of Wolfgang Stegmüller.* Dordrecht: D. Reidel.

Rasmussen, J. (1983). Skill, Rules and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-13, (3)*.

Rasmussen, J., Pejtersen, A. M., & Goodstein, L. P. (1994). *Cognitive Systems engineering*. New York: Wiley.

Reid, G. B., & Nygren, T. E. (1988). The Subjective Workload Assessment Technique: A scaling procedure for measuring Workload. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 185-218). North-Holland, Amsterdam.

Richard, J.-F. (1996). Faut-il revoir la notion de charge mentale ? *Psychologie Française, 41-4*, 309-312.

Richard, J.-F. (2004). *Les activités mentales : de l'interprétation de l'information à l'action.* Paris: Armand Colin.

Ringle, M. D., & Halstead-Nussloch, R. (1989). Shaping user input: a strategy for natural language dialogue design. *Interacting with Computers, 1*(3), 227-244.

Roscoe, A. H. (1987). The practical assessment of pilot workload. In *AGARD-AG-282*. Neuilly Sur Seine: Advisory Group for Aerospace Research and Development.

Roulet, E. (1981). Echanges, interventions et actes de langage dans la structure de la conversation. . *Etudes de linguistique appliquée., 44*, 5-39.

Roulet, E., Auchlin, A., Moeschler, J., Rubattel, C., & Schelling, M. (1985). *L'Articulation du discours en français contemporain.* Berne ; Francfort-sur-Main ; New York: Peter Lang.

Rubio, S., Diaz, E., Martin, J., & Puente, J. M. (2004). Evaluation of Subjective Mental Workload: A comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Applied Psychology, 53(1)*, 61-86.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696-735.

Sadek, D. (1991). *Attitudes mentales et interaction rationnelle : vers une théorie formelle de la communication.* Unpublished thèse, Université Rennes 1, Rennes.

Sadek, D., Bretier, P., & Panaget, F. (1997). ARTIMIS: Natural Dialogue Meets Rational Agency. In *Proceedings of 15th International Joint Conference on Artificial Intelligence* (pp. 1030-1035). Nagoya, Japon.

Saget, S., & Guyomard, M. (2007). Principes d'un modèle collaboratif du dialogue basé sur la notion d'acceptation. In *Annales du LAMSADE, Modèles Formels de l'Interaction (MFI'07), Actes des Quatrièmes Journées Francophones, n°8* (pp. 239-248). Paris, France.

Saussure, F. D. (1913). *Cours de linguistique générale*. Paris: Payot.

Schegloff, E. A. (1990). On the organization of sequences as a source of "coherence" in talk-in-interaction. In B. Dorval (Ed.), *Conversational organization and its development* (pp. 51-77). Norwood: Ablex Publishing.

Schegloff, E. A. (2000). Overlapping Talk and the Organization of Turn-Taking for Conversation. *Language in Society, 29*(1), 1-63.

Schegloff, E. A. (2006). Interaction: The infrastructure for social institutions, the natural ecological niche for language, and the arena in which culture is enacted. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of Human Sociality: Culture, cognition and interaction.* (pp. 70-96). London: Berg.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The Preference for Self-Correction in the Organization of Repair in Conversation. *Language, 53*(2), 361-382.

Schegloff, E. A., & Sacks, H. (1973). Opening Up Closings. *Semiotica, VIII*(4), 289-327.

Schnotz, W. (2005). An Integrated Model of Text and Picture Comprehension. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 49-69). Cambridge: Cambridge University Press.

Schnotz, W., & Kürschner, C. (2007). A Reconsideration of Cognitive Load Theory. *Educational Psychology Review, 19*(4), 469-508.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology, 21*, 211-232.

Schopenhauer, A. (1831). *L'art d'avoir toujours raison*. Paris: Mille et une nuits (traduction de Dominique Miermont).

Seagull, F. J., Wickens, C. D., & Loeb, R. G. (2001). When is Less More? Attention and Workload in Auditory, Visual, and Redundant Patient-Monitoring Conditions. In *Proceedings of the Human Factors and Ergonomics Society 45th annual meeting* (pp. 5).

Searle, J. (1969). *Speech Acts*. Cambridge: Cambridge University Press (version française : (1972) Actes de langage. Paris: Hermann Ed.).

Searle, J., & Vanderveken, D. (1985). *Fundations of The Illocutionary Logic*. UK: Cambridge University Press.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal, 27*, 379-423 and 623-656.

Sheeder, T., & Balogh, J. (2003). Say it like you mean: Priming for structure in caller responses to a spoken dialog system. *International Journal of Speech Technology, 6*, 103-111.

Sinclair, A., & Coulthard, R. M. (1975). *Towards an Analysis of Discourse. The English used by Teachers and Pupils.* Oxford: Oxford University Press.

Sorin, C. (1994). Operational and experimental french telecommunication services using CNET speech recognition and text-to-speech synthesis. In *Paper presented at the 2nd IEEE Workshop on Interactive Voice Technology for Telecommunications Applications (IVTTA)*. Kyoto, Japan.

Spérandio, J.-C. (1972). Charge de travail et régulation des processus opératoires. *Le Travail Humain, 35 (1)*, 85-98.

Spérandio, J.-C. (1977). La régulation des modes opératoires en fonction de la charge de travail chez les contrôleurs de trafic aérien. *Le Travail Humain, 40*, 249-256.

Spérandio, J.-C. (1980). *La psychologie en ergonomie*. Paris: PUF.

Sperber, D., & Wilson, D. (1989). *La Pertinence : Communication et cognition*. Paris: Les éditions de minuit.

Stalnaker, R. C. (1978). Assertion. In P. Cole (Ed.), *Syntax and Semantics.* (Vol. Vol. 9: Pragmatics, pp. 315-332). New-York: Academic Press.

Suchman, L. (1987). *Plans and situated actions: the problem of human/machine communication*. Cambridge: Cambridge University Press.

Sun, Y., Chen, F., Prendinger, H., Chung, V., Shi, Y. D., & Ishizuka, M. (2008). THE HINGE between Input and Output: Understanding the Multimodal Input Fusion Results In an Agent-Based Multimodal Presentation System. In *CHI 2008 Proceedings* (pp. 3483-3488). Florence, Italy.

Sweller, J. (1976). The effect of task complexity and sequence on rule learning and problem solving. *British Journal of Psychology, 67*, 553-558.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12(2)*, 257-285.

Sweller, J. (1999). *Instructional design in technical areas*. Melbourne: ACER Press.

Sweller, J. (2005). Implications of Cognitive Load Theory for Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 19-30). Cambridge: Cambridge University Press.

Sweller, J., & Chandler, P. (1994). Why Some Material Is Difficult to Learn. *Cognition and Instruction, 12 (3)*, 185-233.

Sweller, J., & Cooper, G. A. (1985). The use of worked exemples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*, 59-89.

Sweller, J., & Levine, M. (1982). Effects of goal specificity on means-ends analysis and learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 8*, 463-474.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review, 10 (3)*, 251-296.

Taleb, L. (1996). *Recherche sur l'interaction homme machine : les écarts dans le dialogue informatif finalisé.*, Thèse de l'Université de Paris III (Sorbonne Nouvelle).

Theureau, J. (2002). La notion de charge mentale est-elle soluble dans l'analyse du travail, la conception ergonomique et la recherche neuro-physiologique ? In M. Jourdan & J. Theureau (Eds.), *Charge mentale : Notion floue et vrai problème.* (pp. 41-69). Toulouse: Octarès.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When Two Sensory Modes Are Better Than One. *Journal of Experimental Psychology: Applied, 3 (4)*, 257-287.

Traum, D., & Allen, J. (1991). Causative forces in multi-agent planning. In Y. Demazeau & J. P. Muller (Eds.), *Decentralized A.I.2* (pp. 89-105). Amsterdam: Elsevier Science Publishers B. V.

Tricot, A. (1998). Charge cognitive et apprentissage. Une présentation des travaux de John Sweller. *Revue de Psychologie de l'Education, 1*, 37-64.

Tsang, P. S., & Velasquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics, 39(3)*, 358-381.

Tsang, P. S., & Vidulich, M. A. (2005). Mental workload and situation awareness. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (3rd. ed., Chap. 9 ed., pp. 243-269). New York: Wiley.

Tsang, P. S., & Wilson, G. F. (1997). Mental workload measurement and analysis. . In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 2nd ed., pp. 417-449). New York: Wiley & Sons.

Tsang, P. S., & Wilson, G. F. (2004). Mental workload. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 417-449). New York: Wiley: Taylor and Francis.

van Deemter, K., Krenn, B., Piwek, P., Klesen, M., Schroeder, M., & Baumann, S. (2008). Fully Generated Scripted Dialogue for Embodied Agents. *Artificial Intelligence Journal, 172*(10), 1219-1244.

Van Merriënboer, J. J. G., Schuurman, J. G., de Croock, M. B. M., & Paas, F. G. W. C. (2002). Redirecting learners' attention during training: effects on cognitive load, transfer test, performance and training efficiency. *Learning and Instruction, 12*, 11-37.

Vanderveken, D. (1988). *Les actes de discours. Essai de philosophie du langage et de l'esprit sur la signification des énonciations.* (Pierre Mardaga ed.). Liège, Bruxelles.

Varela, F. J. (1988). *Invitation aux sciences cognitives (Titre original : Cognitive Science : A cartography of Current Ideas)*. Paris: Editions du Seuil.

Varela, F. J., Thompson, E., & Rosch, E. (1993). *L'inscription corporelle de l'esprit - Sciences cognitives et expérience humaine* (V. Havelange, Trans.). Paris: Edition du Seuil.

Vernier, F., & Nigay, L. (2000). A Framework for the Combination and Characterization of Output Modalities. In *Actes de la conférence DSV-IS'2000* (pp. 32-48). Limerick, Ireland.

Vidulich, M. A. (2003). Mental Workload and Situation Awareness: Essential Concepts for Aviation Psychology Practice. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and Practice of Aviation Psychology* (pp. 115-146). London: Lawrence Erlbaum Associates.

Vitense, H. S., Jacko, J. A., & Emery, V. K. (2003). Multimodal feedback: an assessment of performance and mental workload. *Ergonomics, 46 (1-3)*, 68-87.

Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Walker, M. A., Whittaker, S. J., Stentb, A., Maloorc, P., Moored, J., Johnstonc, M., et al. (2004). Generation and evaluation of user tailored responses in multimodal dialogue. *Cognitive Science, 28*, 811-840.

Welford, A. T. (1952). The "psychological refractory period" and the timing of high speed performance - A review and a theory. *British Journal of Psychology, 4 (3)*, 2-19.

Whittaker, S. (2003a). Theories and Methods in Mediated Communication. In A. C. Graesser, M. A. Gernsbacher & S. R. Goldman (Eds.), *Handbook of Discourse Processes* (pp. 243-286). Mahwah, New-Jersey: Lawrence Erlbaum Associates.

Whittaker, S. (2003b). Things to talk about when talking about things. *Human-Computer Interaction, 18(1-2)*, 149-170.

Whittaker, S., Brennan, S. E., & Clark, H. H. (1991). Coordinating activity: an analysis of interaction in computer supported cooperative work. In *CHI* (pp. 1-8).

Whittaker, S., & Walker, M. A. (1991). Towards a Theory of Multimodal Interaction. In M. T. Maybury (Ed.), *Working Notes from AAAI Workshop on Intelligent Multimedia Interfaces, Ninth NCAI*. Anaheim, CA: AAAI Press.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63-102). New-York, NY: Academic Press.

Wickens, C. D. (1987). Information processing, decision making, and cognition. In G. Salvendy (Ed.), *Handbook of Human Factors* (pp. 72-107). New-York: John Wiley & Sons.

Wickens, C. D. (2000). Human Factors in Vector Map Design: The Importance of Task-Display Dependence. *Journal of Navigation, 53*, 54-67.

Wickens, C. D. (2001). Wokload and Situation Awareness. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload and fatigue: theory, research and practice* (pp. 443-450). New Jersey: Lawrence Erlbaum.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoritical Issues in Ergonomics Science, 3(2)*, 159-177.

Wickens, C. D., Goh, J., Helleberg, J. R., & Talleur, D. A. (2002). *Modality Differences in Advanced Cockpit Displays: Comparing Auditory Vision and Redundancy for Navigational Communications and Traffic Awareness* (Technical report No. ARL-02-8/NASA-02-6). Savoy, Illinois: Aviation Research Lab: Institute of Aviation.

Wickens, C. D., McCarley, J. S., Alexander, A. L., Thomas, L. C., Ambinder, M., & Zheng, S. (2005). *Attention-Situation Awareness (A-SA) Model of Pilot Error* (Rapport technique No. AHFD-04-15 (contract NASA NAG 2-1535)). Moffett Field, CA: NASA Ames Research Center.

Wickens, C. D., & Seppelt, B. (2002). *Interference With Driving or In-Vehicle Task Information: The Effects of Auditory Versus Visual Delivery* (Technical report No. AHFD-02-18/GM-02-3). Savoy, Illinois: Aviation Human Factors Division: Institute of Aviation.

Wiener, N. (1948). *Cybernetics: Control and communication in the animal and the machine*. Cambridge, MA: MIT Press.

Wilamowitz-Moellendorff, M., Müller, C., Jameson, A., Brandherm, B., & Schwartz, T. (2005). Recognition of Time Pressure via Physiological Sensors: Is the User's Motion a Help or a Hindrance? In *Proceedings of the Workshop on Adapting the Interaction Style to*

Affective Factors in conjunction with the User Modeling (UM05) (pp. 43-48). Edinburgh, UK.

Wilkie, J., Jack, M. A., & Littlewood, P. J. (2005). System-initiated digressive proposals in automated human-computer telephone dialogues: the use of contrasting politeness strategies. *International Journal of Human-Computer Studies, 62*, 41-71.

Wilson, T. D. (1997). Information behavior: An interdisciplinary perspective. *Information Processing & Management, 33 (4)*, 551-572.

Wittgenstein, L. (1922). *Tractatus logico-philosophicus* (Gallimard ed.). Paris: Traduction de Gilles-Gaston Granger.

Wittgenstein, L. (1953). *Recherches philosophiques*: Gallimard.

Wooldridge, M. (2002). *An Introduction to MultiAgent Systems*. Liverpool, UK: John Wiley & Sons Ltd.

Yeung, A. S., Lee, C. F. K., Pena, I. M., & Ryde, J. (2000). *Toward a Subjective Mental Workload Measure.* Paper presented at the International Congress for School Effectiveness and Improvement, Hong-Kong.

Yin, B., & Fang, C. (2007). Towards Automatic Cognitive Load Measurement from Speech Analysis. In *12th HCII* (pp. 1011-1020). Beijing (China): Springer-Verlag.

Zoltan-Ford, E. (1991). How to get people to say and type what computers can understand. *International Journal of Man-Machine Studies, 34*, 527-547.

Zouinar, M. (2000). *Contribution à l'étude de la coopération et du partage d'informations contextuelles dans les environnements de travail complexes.* Unpublished Thèse, Conservatoire National des Arts et Métiers (CNAM), Paris.

Zue, V., Seneff, S., Glass, J., Polifroni, J., Pao, C., Hazent, J., et al. (2000). JUPITER : A telephone-based conversational interface for weather information. In *IEEE Transactions On Speech and Audio Processing.* (Vol. Vol. X, pp. 100-112).

ANNEXES

8.4 Cahier de consigne des expériences 1 et 4

- ***Page de garde***

Cette page présente la consigne générale

Merci d'avoir accepté de participer à cette étude

Vous allez participer à une recherche qui est focalisée sur l'interaction avec un service de prise de rendez-vous. Pour cette étude, vous chercherez à prendre un rendez-vous avec un médecin dans un hôpital.

L'application que vous allez tester est basée sur un logiciel de reconnaissance vocale. Vous n'avez pas besoin du clavier ni de la souris pour la commander, mais uniquement de lui parler. Ce logiciel comprend le langage naturel et vous n'avez pas besoin de modifier votre façon de parler habituelle.

Vous allez faire 3 appels consécutifs sur le service. Suite à chaque appel vous répondrez à un questionnaire dont le but est d'évaluer votre charge mentale pendant l'utilisation du service.

Nous sommes le mercredi 31 août 2005.
Vous êtes un patient de l'hôpital.
Vous voulez prendre un rendez-vous avec un médecin.

Les pages suivantes de ce livret indiquent les horaires auxquels vous êtes disponibles et le nom du médecin que vous désirez rencontrer.
Comme en situation réelle, d'autres patients ont déjà pris rendez-vous avec ce médecin. Votre objectif est de chercher à obtenir un rendez-vous qui respecte les contraintes de votre emploi du temps. Comportez vous comme si vous étiez réellement chez vous.

D'un appel à l'autre, ne tenez pas compte des rendez-vous pris à l'appel précédent.

- ***Scénario 1***

Consigne donnée aux participants pour le premier appel auprès du service expérimental.

Appel 1 : Vous avez besoin d'un rendez-vous chez le Dr Dubois.

Voici votre emploi du temps du 1er au 7 septembre :

	Aujourd'hui						
	Mercredi 31	Jeudi 1er	Vendredi 2	Samedi 3	Lundi 5	Mardi 6	Mercredi 7
8H - 8H30							
8H30 - 9H							
9H - 9H30							
9H30 - 10H							
10H - 10H30							
10H30 - 11H							
11H - 11H30							
11H30 - 12H							
12H - 12H30							
14H - 14H30							
14H30 - 15H							
15H - 15H30							
15H30 - 16H							
16H - 16H30							
16H30 -17H							
17H - 17H30							
17H30 - 18H							
18H - 18H30							

Vous êtes indisponible sur ces plages

Plages disponibles

- ***Scénario 2***

Consigne donnée aux participants pour le deuxième appel auprès du service expérimental.

Appel 2 : Vous envisagez d'aller voir le Dr Latour.

Voici votre emploi du temps du 1er au 7 septembre :

	Aujourd'hui						
	Mercredi 31	Jeudi 1er	Vendredi 2	Samedi 3	Lundi 5	Mardi 6	Mercredi 7
8H - 8H30							
8H30 - 9H							
9H - 9H30							
9H30 - 10H							
10H - 10H30							
10H30 - 11H							
11H - 11H30							
11H30 - 12H							
12H - 12H30							
14H - 14H30							
14H30 - 15H							
15H - 15H30							
15H30 - 16H							
16H - 16H30							
16H30 -17H							
17H - 17H30							
17H30 - 18H							
18H - 18H30							

Vous êtes indisponible sur ces plages

Plages disponibles

- ***Scénario 3***

Consigne donnée aux participants pour le troisième appel auprès du service expérimental.

Appel 3 : Vous souhaitez prendre un rendez-vous avec le Dr Lebrun.

Voici votre emploi du temps du 1er au 7 septembre :

	Aujourd'hui						
	Mercredi 31	Jeudi 1er	Vendredi 2	Samedi 3	Lundi 5	Mardi 6	Mercredi 7
8H - 8H30							
8H30 - 9H							
9H - 9H30							
9H30 - 10H							
10H - 10H30							
10H30 - 11H							
11H - 11H30							
11H30 - 12H							
12H - 12H30							
14H - 14H30							
14H30 - 15H							
15H - 15H30							
15H30 - 16H							
16H - 16H30							
16H30 -17H							
17H - 17H30							
17H30 - 18H							
18H - 18H30							

Vous êtes indisponible sur ces plages

Plages disponibles

8.5 Questionnaires d'évaluation de la charge cognitive

8.5.1 NASA-TLX

La version du questionnaire NASA-TLX utilisée dans les expériences 1, 3 et 4 était une version française du questionnaire proposé par Hart & Staveland (1988). Les formulations utilisées correspondaient à une traduction littérale de la version d'origine proposée par ces auteurs.

- ***Consigne générale***

La consigne générale était présentée au participant en début d'appel :

Nous sommes intéressés par vos expériences au cours des différentes conditions de la tâche. Voici la procédure qui va être utilisée pour examiner vos expériences.

D'une manière générale nous examinons la charge de travail. Les facteurs qui influencent la charge de travail peuvent être liés à la tâche elle-même, votre sentiment envers votre propre performance, l'effort que vous avez fourni ou la frustration que vous avez ressenti.

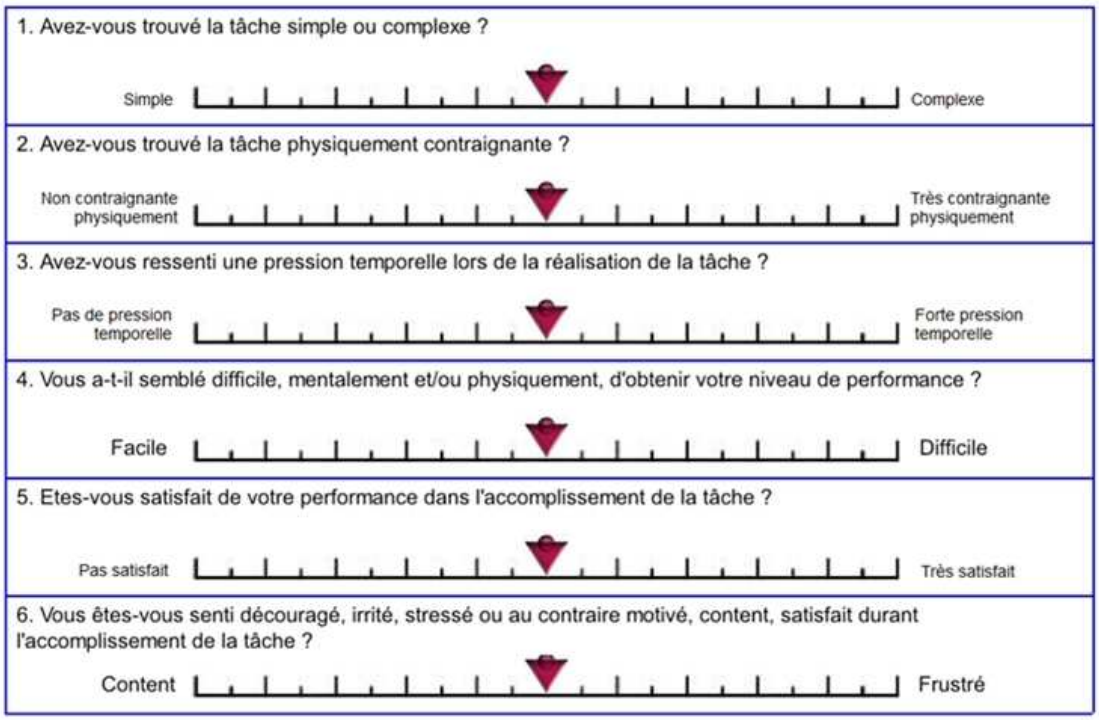
Nous vous demandons d'évaluer individuellement ces dimensions de la charge de travail. Si vous avez la moindre question sur l'une de ces dimensions, n'hésitez pas à questionner votre expérimentateur.

Après avoir réalisé la tâche, 6 échelles d'évaluation vous seront présentées. Vous évaluerez la tâche en déplaçant le marqueur sur l'échelle à l'endroit qui correspond le mieux à votre expérience. Chaque ligne est définie par 2 descripteurs placés à ses extrêmes. Pour répondre, veillez à tenir compte des conditions de votre tâche et à considérer individuellement chacune des échelles.

Votre évaluation aura un rôle important dans l'étude qui est conduite, et votre participation active est essentielle au succès de cette expérience, et elle est grandement appréciée.

- ***Evaluation des dimensions***

Les évaluations étaient faites sur des échelles continues. La graduation comportait 20 points, comme indiqué par Hart & Staveland (1988). Les valeurs enregistrées étaient comprises entre 0 et 100 (chaque graduation comprenait 5 valeurs possibles). Les mesures étaient donc plus précises que ce que recommandent les auteurs.



- ***Procédure de comparaison par paires (« Pair-wised comparison »)***

La procédure de comparaison par paires commençait par la consigne ci-dessous.

Le poids des dimensions qui composent la charge de travail diffère d'une tâche à l'autre. Par exemple, une tâche peut-être difficile parce qu'elle doit être réalisée dans un temps court, une autre parce qu'elle demande un effort mental important.

Vous allez maintenant évaluer l'importance relative des 6 dimensions pour la tâche que vous avez réalisée.

Des affirmations conformes aux différentes échelles vont vous être présentées par paires. Vous devez simplement choisir celle qui correspond le mieux à l'expérience que vous avez eu au cours de la réalisation de la tâche.

Lorsque vous avez fait votre choix cliquez sur le texte de l'affirmation que vous avez choisi.

Si cette procédure n'est pas claire, n'hésitez pas à questionner l'expérimentateur.

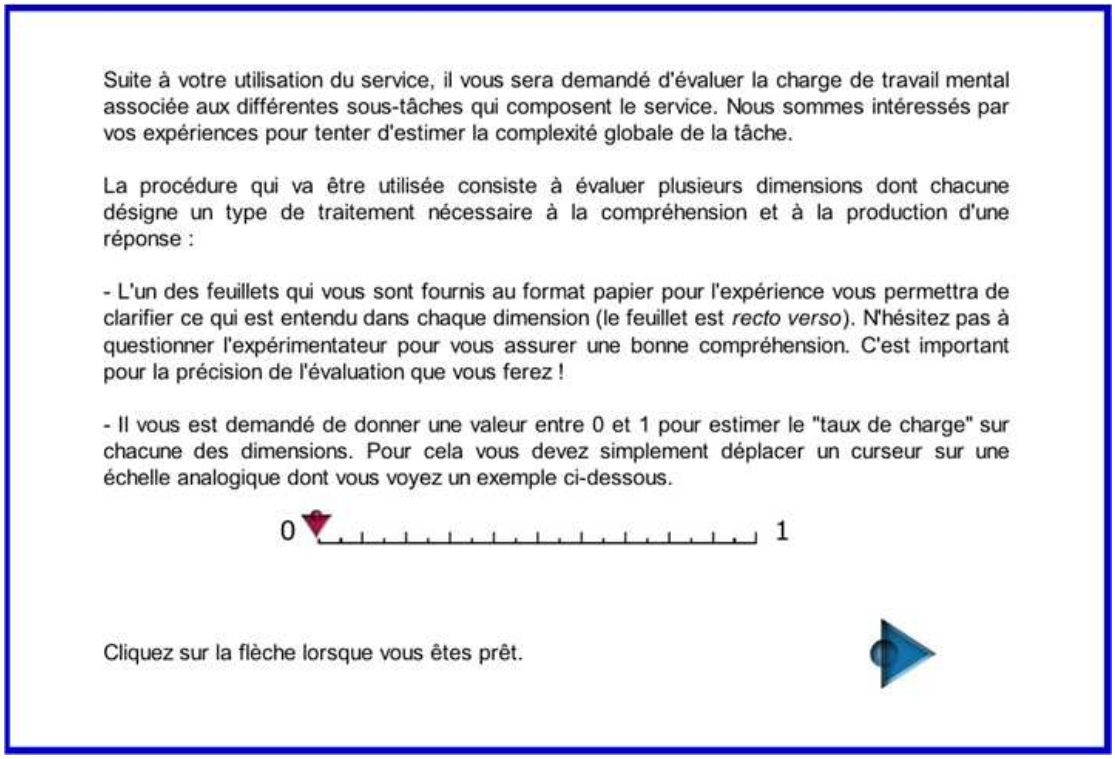
Les affirmations qui devaient être sélectionnées sont listées ci-dessous :

a. *« La tâche était complexe »*
b. *« La tâche était physiquement contraignante »*
c. *« Le temps imparti était réduit pour la tâche »*
d. *« Ma performance a été difficile à obtenir »*
e. *« Ma performance est satisfaisante »*
f. *« L'épreuve était désagréable »*

8.5.2 Workload Profile

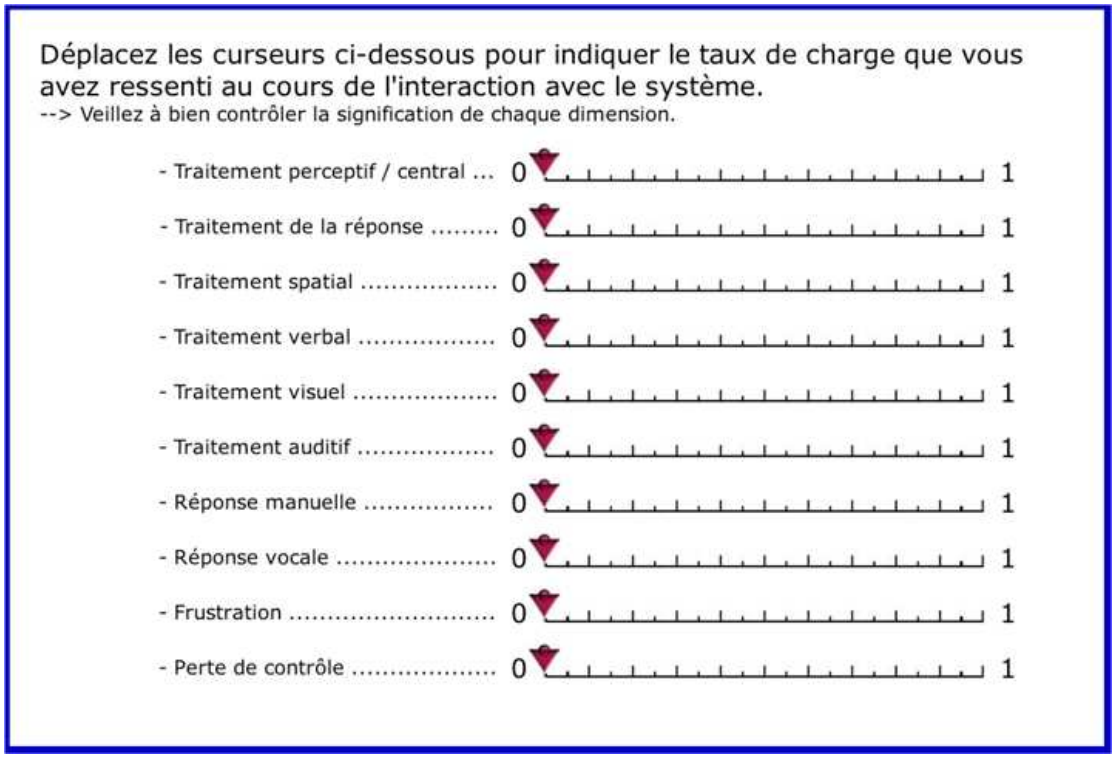
- ***Consigne générale***

La consigne générale était présentée au participant en début d'expérience :



- ***Evaluation des dimensions***

Les dimensions étaient évaluées après chaque appel :



Les mesures faites sur des échelles graduées en 20 points. Les valeurs enregistrées étaient comprises entre 0 et 100, ce qui correspond intuitivement à un « taux de charge ».

- ***Fiche de présentation des dimensions***

La fiche suivante était remise au participant (version papier). C'est une traduction en français de la version proposée par Tsang et Velasquez (1996). Elle comprend 8 dimensions.

Dimensions de la charge de travail mental

1. Niveau de traitement

(1) Traitement perceptif / central

Ce sont les ressources attentionnelles requises pour des activités comme la perception (détecter, reconnaître et identifier des objets), la mémorisation, la résolution de problème et la prise de décision.

(2) Traitement de la réponse

Ce sont les ressources attentionnelles requises pour des activités comme la sélection de réponse et l'exécution. Par exemple, il y a 3 pédales dans une automobile standard ; pour arrêter l'automobile, il est nécessaire de choisir et d'appuyer sur la pédale appropriée.

2. Code de traitement

(3) Traitement spatial

Certaines tâches sont spatiales par nature. La conduite, par exemple, oblige à porter son attention sur la position de la voiture, la distance entre la position actuelle et la prochaine indication de stop, la direction que prend la voiture, etc.

(4) Traitement verbal

D'autres tâches sont verbales par nature. Par exemple, la lecture implique surtout le traitement de matériel verbal, linguistique.

3. Modalité perceptive

(5) Traitement visuel

L'exécution de certaines tâches est basée sur les informations visuelles reçues. Par exemple, pour jouer au basket il est nécessaire de contrôler visuellement l'emplacement physique et la vitesse de déplacement du ballon. Regarde la télévision est un autre exemple de tâche qui requière des ressources visuelles.

(6) Traitement auditif

Pour d'autres tâches l'exécution est basée sur les informations auditives. Par exemple, écouter un enseignant lors d'un cours est une tâche qui requière de l'attention auditive. Ecouter de la musique est un autre exemple.

Notez que des informations spatiales peuvent être traitées visuellement ou auditivement. Par exemple, vous pouvez vous rendre dans un nouveau restaurant en suivant une carte (traitement visuel) ou en suivant les indications d'un ami (traitement auditif). De façon similaire, une information verbale peut être traitée visuellement ou auditivement. Ecouter les informations à la radio nécessite un traitement auditif du matériel verbal ; lire les informations dans un journal nécessite un traitement visuel du matériel verbal.

4. Modalité d'action

(7) Réponse manuelle

Certaines tâches impliquent une attention considérable pour produire la réponse manuelle comme la saisie de texte ou jouer du piano.

(8) Réponse vocale

D'autres tâches impliquent plutôt des réponses vocales. Par exemple, s'engager dans une conversation demande de faire attention à la production de la parole lors des réponses.

Les dimensions (9) et (10) ont été ajoutées pour prendre en compte des valeurs de nature émotionnelle. Ces deux dimensions sont inspirées du modèle de Lazarus et Folkman (1984).

5. Niveau émotionnel

(9) Frustration

Certaines tâches vous donnent le sentiment d'accomplir pleinement vos objectifs alors que d'autres peuvent au contraire vous sembler aller à l'encontre de ce que vous tentez de faire.

(10) Perte de contrôle

Dans certains cas, vous pouvez contrôler totalement le déroulement des événements, alors que dans d'autres cas vous pouvez avoir le sentiment de perdre la main.

TABLE DES MATIERES DETAILLEE

LISTE DES TABLEAUX

LISTE DES FIGURES

Eléments pour la conception d'énoncés multimodaux en Dialogue Homme Machine –
Pourquoi l'unité d'analyse psychologique est l'Action et non l'Information

Cette thèse vise l'amélioration des capacités de communication des systèmes de *Dialogue Homme Machine*. Les travaux présentés visent à analyser les actions du système et leurs effets dans le dialogue. Cette *problématique de conception des actes du système* suppose de disposer d'outils conceptuels propices à l'analyse et renvoie à la problématique du *fonctionnement cognitif de l'individu humain* dans la communication. La partie théorique pose cette problématique appliquée dans un contexte interdisciplinaire (entre *linguistique*, *ingénierie* et *psychologie*). Cette présentation permet d'opposer deux points de vue dans la partie expérimentale : (1) l'*approche pragmatique*, qui analyse les *actes des partenaires*, selon un point de vue sociocognitif, et (2) l'*approche cognitive*, qui analyse les *processus de traitement de l'information*, centrée sur le niveau individuel.

Cinq expériences (protocole du *Magicien d'Oz*) sont présentées. Dans les deux premières, le système communiquait en mode vocal (*énoncés auditifs*). Ces deux expériences mettent en évidence l'utilité et les effets des *messages d'aide* (*aides procédurales*) et de la *syntaxe* dans les énoncés auditifs. Dans les trois expériences suivantes, le système communiquait en mode multimodal (*énoncés audio-visuels*). Une *catégorisation des types d'information à présenter* a été introduite. Une *règle d'attribution* des modes de présentation (*auditif* et/ou *visuel*) aux différents types d'information (*'écho'*, *'réponse'*, *'relance'*) a été proposée pour concevoir des *'stratégies de présentation'* innovantes. Ces trois expériences ont permis de démontrer l'intérêt des principes d'analyse utilisés pour la *conception des stratégies de présentation*. Elles montrent l'importance de la *relation type-mode* pour prédire les effets des actes. Les résultats obtenus permettent de valider l'*approche pragmatique* contre l'*approche cognitive*. La discussion permet d'aborder les implications de ces résultats, sous l'angle de la conception des énoncés des systèmes de DHM et sous l'angle des conséquences théoriques qui peuvent être tirées de ces résultats.

Mots-clés : Dialogue Homme Machine ; Multimodalité ; Conception d'énoncés ; Charge cognitive ; Action

Elements for multimodal utterances design in Human-Computer Dialogue –
Why psychological analysis unit is Action and not Information

This thesis aims to improve communication skills of *Human Computer Dialogue* systems through the analysis of system actions and their effects in dialogue. This problematic supposes the use of adapted conceptual tools for the analysis and is linked to the problem of *individual human cognitive processing* during communication. The theoretical part explores this applied problematic in an interdisciplinary context (between *linguistics, engineering and psychology*). This presentation allows opposing two different viewpoints in the experimental part: (1) *pragmatic approach* analyses the *partners acts*, following a socio-cognitive viewpoint and (2) *cognitive approach* analyses the *information processing*, focalized on the individual level.

Five experiments (*Wizard of Oz* protocol) are presented. In the two firsts system was communicating in spoken mode (*auditory utterances*). These two experiments gave evidence for the utility and effects of *help messages* (procedural content) and *syntax* in auditory utterances. In the three last experiments, the system was using multimodal communication (*audio-visual utterances*). A categorization of information in three types has been introduced (*'echo'*, *'response'* and *'opening'*). A rule was proposed to attribute presentation modes (*auditory* and/or *visual*) in function of information types, in order to design innovative *'presentation strategies'*. These three experiments demonstrated the relevance of the analysis principles used to design presentation strategies. They showed *type-mode relation* importance to predict the effects of acts. Obtained results allow validating the *pragmatic approach* against the *cognitive one*. Implications of these results are discussed in the conclusion part, first for the design of system utterances in HCD and, secondly about theoretical consequences to consider.

Keywords: Human Computer Dialogue; Multimodality; Utterances design; Cognitive load; Action