
UNIVERSITE DE CAEN

INSTITUT DES SCIENCES DE LA MATIERE
ET DU RAYONNEMENT

GROUPE DE RECHERCHES EN IMAGE, INSTRUMENTATION,
INFORMATIQUE DE CAEN

Agents informatiques interactifs

Applications linguistiques et biomédicales

Dossier d'habilitation à diriger des recherches
présenté à l'Université de Caen
le 29 octobre 1997

par

Pierre NUGUES

devant un jury composé de

Patrice Enjalbert, Professeur
Jean-Paul Haton, Professeur
Jean-Marie Pierrel, Professeur
Gérard Sabah, Directeur de recherche
Bernard Victorri, Directeur de recherche

ANNEE 1996-1997

Remerciements

Je remercie très vivement et très sincèrement :

Bernard Victorri, Directeur de Recherches au CNRS, pour m'avoir dirigé pendant la rédaction de cette habilitation, pour ses conseils, sa patience et sa gentillesse.

Jean-Marie Pierrel, Professeur à l'Université Henri Poincaré, et Gérard Sabah, Directeur de Recherches au CNRS, qui ont accepté d'être les rapporteurs de ce document et de juger mon travail.

Patrice Enjalbert, Professeur à l'Université de Caen, et Jean-Paul Haton, Professeur à l'Université Henri Poincaré, qui ont bien voulu être membres du jury.

À Charlotte, ma bien aimée

1.
Présentation générale

Les recherches scientifiques que j'ai menées relèvent, pour une large part, du génie biomédical et de l'informatique linguistique. Je me suis consacré au premier thème au cours d'une thèse de doctorat à l'université de Nancy (1989) en construisant un système d'analyse et d'interprétation d'images. Je me suis ensuite tourné vers le traitement automatique de la langue et les agents conversationnels après avoir obtenu un poste de Maître de conférences à l'Institut des Sciences de la Matière et du Rayonnement (ISMRA) de Caen.

À l'ISMRA, j'ai entamé un nouveau domaine de recherches, centré sur l'analyse syntaxique et sémantique ainsi que sur le dialogue et l'interaction multimédias. Les nouveaux systèmes que j'ai réalisés ont largement bénéficié des concepts et des méthodes qui m'étaient déjà familiers. Indirectement, ils m'ont ainsi donné l'occasion de poursuivre et d'approfondir dans la voie ouverte par la thèse en même temps que je renouvelais mon champ d'étude.

Je vais m'efforcer dans cette présentation sous forme chronologique de retracer les étapes de mon parcours et de montrer les fils conducteurs qui les relient.

1.1 Grenoble

J'ai réalisé mon premier projet scientifique au Laboratoire d'Électronique et des Technologies Informatiques (LÉTI) du CEA, à Grenoble, sous la direction de Philippe Garderet, en 1984. Il a eu lieu à l'occasion de mon stage de fin d'études d'ingénieur. Philippe Garderet dirigeait alors un service spécialisé dans le traitement d'images – il avait à son actif la réalisation du premier scanner tomographe français. À l'époque, Philippe entamait une nouvelle activité et il pilotait le développement d'un analyseur d'images d'électrophorèses bidimensionnelles.

Les électrophorèses bidimensionnelles sont des analyses biochimiques. Elles permettent de connaître la composition en protéines d'un tissu en les répartissant sur un gel. Chaque protéine du tissu apparaît à des coordonnées caractéristiques. Les protéines prennent l'aspect d'amas noirs sur la plaque de gel. L'ensemble ressemble à un buvard taché.

L'équipe du LÉTI utilisait des images numériques de gels provenant du laboratoire Argonne (ANL) à Chicago aux États-Unis. Elle avait réalisé un programme qui détectait le sommet des taches. Mon travail a consisté à modéliser chacune des taches sous la forme d'une cloche gaussienne bidimensionnelle et à écrire l'algorithme d'approximation de ses paramètres.

Les approximations publiées dans la littérature scientifique faisaient appel à une méthode de moindres carrés non linéaires. J'ai tout d'abord programmé cet algorithme numérique. Après quelques essais, je me suis aperçu qu'il demandait des calculs importants dès que les amas de protéines connexes étaient nombreux. Nous disposions d'un ordinateur VAX pour tout le service et la méthode se révélait, en fait, très longue dès qu'il y avait plus de deux taches dans l'amas. Les laboratoires américains concurrents de l'époque utilisaient des processeurs vectoriels.

Dans la suite du stage, j'ai simplifié cette méthode pour l'appliquer, non pas sur un amas en entier mais itérativement, sur chacun des sommets de l'amas. La fenêtre d'analyse était alors réduite et centrée sur un extremum. Le problème devenait traitable dans un temps raisonnable. De cette façon, j'ai pu obtenir une approximation relativement bonne des paramètres de chacune des taches d'une image.

1.2 Nancy

À la suite du stage de Grenoble, j'ai cherché à poursuivre dans la voie de ce que j'avais entrepris. Elle avait toujours pour centre les images d'électrophorèses. Ce sujet se trouvait à la rencontre de mes connaissances techniques du moment – l'informatique et en particulier le traitement d'images et le calcul numérique – et de l'analyse génétique. C'était ce domaine qui me passionnait.

L'idée était de relier automatiquement, grâce à un système d'interprétation, les résultats de cette analyse biochimique à un état clinique ou génétique. Après quelques démarches, la société Cognitech a accepté et financé une thèse de doctorat sous la forme d'une convention CIFRE. J'ai travaillé à cette thèse pendant deux ans et demi sous la direction de Jean-Paul Haton au Centre de Recherche en Informatique de Nancy (CRIN) et j'ai réalisé un système d'interprétation d'électrophorèses bidimensionnelles. La durée de la convention a aussi recouvert le stage de six mois de mon DEA à l'université Pierre-et-Marie-Curie (1986). Pendant ce stage, j'ai travaillé à un système de segmentation et de lecture automatique d'annonces d'offres d'emploi.

L'interprétation d'une image de gel consiste à localiser les protéines, à les identifier et à les relier à une connaissance clinique. La difficulté provient de la variabilité des images pour un même tissu et de la dégradation éventuelle du gel. Ceci tient essentiellement au procédé biochimique – il s'est sensiblement amélioré depuis. Durant ma thèse, j'ai mis en œuvre un système qui permet de nettoyer les images, de détecter les taches de protéine et de les recalculer les unes par rapport aux autres. Le processus conduit à identifier les taches à des protéines connues ainsi que des groupes de protéines d'une même famille.

Pour réaliser le système d'interprétation, je me fondais sur des connaissances en traitement d'images. Elles devaient améliorer l'aspect des gels et détecter les taches. J'avais eu l'occasion d'aborder ces thèmes à Grenoble. Je devais aussi y intégrer des techniques d'intelligence artificielle¹. Ces dernières auraient pour tâche d'apparier des modèles géométriques sur l'image, d'identifier les points, de reconnaître les configurations...

L'équipe de Jean-Paul Haton réalisait la fusion de ces compétences : par les chercheurs qui l'entouraient et par les techniques qu'elle voulait mettre en œuvre. Son activité au CRIN s'appelait alors *Reconnaissance des formes et Intelligence artificielle*.

En 1986, beaucoup de laboratoires ayant une activité en intelligence artificielle tentaient de résoudre des problèmes semblables. Ils se préoccupaient de la meilleure

¹ Ce terme a suscité des controverses (Searle 1980) et il est peut être mal approprié. Cependant, il est difficile de lui trouver un remplaçant.

façon d'intégrer des connaissances « hétérogènes » et des méthodes de traitement dans une architecture unique. Les options étaient de piloter les différents modules avec plus ou moins de souplesse et d'autonomie pour chacune des entités. Suivant ces options, les architectures prenaient le nom de « blackboard » ou de systèmes multi-agents.

Lors de la première partie de ma thèse, j'ai conçu et implanté les opérateurs de traitement d'images ainsi que les modules d'identification et d'appariement nécessaires à l'interprétation des gels. Je les ai intégrés dans une architecture modulaire qui relevait du modèle multi-agents. Le système se destinait à identifier une classe particulière de protéines du plasma : les apolipoprotéines. Il a pu produire des identifications exactes sur les gels de tests à l'exception d'un gel très dégradé.

L'interprétation des gels par le système que j'ai conçu fait appel à une connaissance préexistante sur les protéines en jeu : où elles se trouvent sur le gel, quelle est leur influence. Si certaines protéines sont connues, notamment pour des liquides biologiques tels que le plasma, beaucoup d'autres n'ont même pas de nom. Or, un gel produit une quantité considérable de données ; couramment plusieurs milliers de taches. Il me semblait difficile d'envisager de les explorer toutes à la main. Par exemple, pour savoir quelles sont les protéines qui réagissent à certains composés ou bien si des groupes de protéines réagissent ensemble.

Dans la seconde partie de ma thèse, j'ai cherché à identifier automatiquement des protéines liées à un état génétique ou clinique, sans en avoir une connaissance *a priori*. Pour ceci, j'ai conçu et programmé des méthodes symboliques pour analyser les résultats d'électrophorèses de plusieurs types de cellules. Les groupes de gels correspondaient à des étapes de la croissance du muscle et la quasi totalité des protéines qui composaient ces gels étaient inconnues. Le but était d'identifier quels étaient les paramètres – les protéines – qui « expliquaient » le mieux un état particulier.

Les méthodes que j'ai mises en œuvre tenaient de l'apprentissage automatique et plus particulièrement de la classification conceptuelle. L'idée sous-jacente était qu'une démarche symbolique intégrerait mieux les connaissances du domaine que des méthodes purement numériques. Ces connaissances étant nécessaires pour guider la classification.

Avec des données fournies par Robert Whalen de l'Institut Pasteur, j'ai pu produire, pour chacun des types de cellules, des formules protéiques discriminantes. Un type de cellule se caractérisant par la présence ou l'absence de certaines protéines ou par des combinaisons de protéines. L'expérience n'a cependant pas été menée jusqu'au bout. L'exploration biologique des protéines identifiées resterait à faire. On peut imaginer que certaines protéines expliquent un état physiologique particulier (Nugues et al. 1989).

J'ai terminé mon travail à Nancy en décembre 1988 et j'ai soutenu ma thèse en mars 1989.

1.3 Caen

Après ma thèse, j'ai travaillé pendant un an pour la société Cognitech, notamment dans le cadre d'un projet avec l'Aérospatiale. Il avait comme objectif de recalculer automatiquement des images de radar sur une carte géographique. Dans un contexte très différent, le problème était assez semblable à celui que j'avais eu à résoudre avec les électrophorèses. Il fallait mettre en œuvre des opérateurs de traitement d'images pour extraire des paramètres caractéristiques, en construire une représentation et l'apparier à une zone précise de la carte.

Dans le projet pour l'Aérospatiale, j'ai utilisé des techniques relativement classiques de traitement d'images. La transformée de Hough-Radon, notamment, a servi pour détecter les routes de l'image radar. Le recalage des routes sur la carte a fait appel à des méthodes de raisonnement hypothétique. Pour ceci, j'ai utilisé un système de maintien de la vérité fondé sur les hypothèses (*Assumption-Based Truth Maintenance System*) (de Kleer 1986).

Le raisonnement avec un ATMS permet d'explorer plusieurs hypothèses en parallèle en ajoutant des faits par incréments. Si un fait rentre en contradiction avec une hypothèse, cette dernière est supprimée alors qu'on continue à développer les autres. Ce mécanisme de raisonnement a l'inconvénient d'être coûteux. En revanche, il semble mieux fondé du point de vue théorique que les autres systèmes à bases de connaissances. Il a permis de recalculer trois des quatre images de test qui m'avaient été données à traiter.

En octobre 1990, j'ai obtenu un poste de Maître de Conférences en informatique à l'ISMRA. Je me suis alors déplacé vers le domaine de l'informatique linguistique et du dialogue oral avec une machine. J'ai ainsi changé de sujet, tout en restant dans un premier temps dans des applications biomédicales.

De 1991 à 1993, j'ai dirigé la construction d'un prototype de dictée interactive de comptes rendus médicaux. Ce projet a recouvert deux stages de DEA et des projets d'élèves ingénieur. Le prototype de dictée – DictaMed – s'est inspiré au départ des idées et des travaux de Jean-Marie Pierrel (1987) du CRIN. Il permettait de dicter des comptes rendus médicaux à une machine puis de la questionner.

DictaMed comportait un dispositif externe de reconnaissance vocale de mots isolés. Il effectuait l'analyse syntaxique des mots et des phrases. Il créait les objets sémantiques correspondants en même temps que la dictée. Le système interagissait avec l'utilisateur lorsqu'il détectait des incohérences ou pour demander des précisions. On pouvait ensuite questionner la machine sur ce que l'utilisateur avait mentionné et elle synthétisait oralement les réponses.

Bien que le domaine ait été complètement différent, l'architecture était voisine de celle du système que j'avais construit pendant ma thèse. Nous avons séparé la syntaxe, la sémantique et le dialogue en modules indépendants. L'ensemble étant assemblé comme un système multi-agents centré sur un noyau-objet. DictaMed s'incorporait dans un environnement graphique qui permettait les commandes au clavier ou à la souris. Il comportait des fenêtres graphiques de messages pour guider l'utilisateur.

Le noyau-objet formait le cœur du système. Il modélisait toutes les entités syntaxiques et sémantiques. Ces dernières correspondaient à des concepts médicaux attachés à un examen cardiaque (l'analyse de Holter tensionnel). En contraste avec

mon travail de thèse, le noyau-objet permettait une modélisation universelle et partageable par tous les modules des entités du système.

La grammaire utilisait des règles syntagmatiques écrites avec la notation DCG (Pereira & Warren 1980). Elles comportaient des équations d'unification. Ces équations nous permettaient notamment de modéliser la valence des verbes. Chaque verbe du lexique indiquait la liste de ses compléments possibles. Les compléments se classaient en obligatoires, facultatifs ou interdits. La prise en compte de ce contexte permettait de restreindre le nombre de structures possibles lors de l'analyse syntaxique.

L'analyse sémantique créait ou référençait les objets médicaux correspondant aux entités mentionnées dans la phrase. Dans les corpus d'exemples que nous avons, il arrivait que le médecin émette une hypothèse puis revienne dessus et la contredise. Pour gérer la cohérence logique du compte rendu, DictaMed créait les objets avec un système de maintien de la vérité. Le mécanisme était semblable à celui utilisé pour l'Aérospatiale. Cependant, DictaMed ne réalisait pas le raisonnement. Il suivait simplement celui du médecin. Pour cette raison, nous avons choisi d'implanter un TMS (*Truth Maintenance System*) (Doyle 1979). Le TMS effectuait des vérifications de cohérence sur les attributs des objets et sur l'enchaînement des thèmes de la dictée. Une fois la valeur d'un attribut donnée, il propageait les contraintes correspondantes. Si l'utilisateur retirait cette valeur, le système retirait aussi les contraintes qu'il avait posées.

Enfin, un automate pilotait le dialogue. Il gérait les tours de parole et avertissait des conflits. Le système pouvait répondre à des questions sur les valeurs des attributs et sur le déroulement de la dictée.

À partir de 1993, j'ai eu la responsabilité d'encadrer une première thèse, puis en 1994, une deuxième. Ces thèses, ainsi que deux stages de DEA, en 1995 et 1996, et le *Diplomarbeit* de deux étudiants allemands, m'ont permis de réaliser un projet plus ambitieux. Mon travail est à ce moment devenu celui d'une équipe que j'ai dirigé dans sa totalité.

Les compétences acquises avec DictaMed, les démonstrations du système et les publications associées, m'ont permis, à peu près au même moment, de faire partie d'un groupe de travail européen (COST-14). Ce groupe a pris la forme d'une collaboration entre plusieurs centres de recherches universitaires ou industriels. Parmi eux, il y avait l'Université de Nottingham, le Swedish Institute of Computer Science, l'Université de Lancaster, le Gesellschaft für Mathematik und Datenverarbeitung. Le thème de l'étude était le travail en groupe à l'aide de l'ordinateur (*Computer Supported Cooperative Work*). Le groupe s'était fixé comme objectif de produire des maquettes d'outils de téléconférence et de travail en groupe. Ces maquettes devaient exploiter des techniques de réalité virtuelle ou « améliorée » (*augmented*).

Dans un environnement virtuel de travail en groupe, les espaces de travail sont modélisés par des scènes tridimensionnelles. Les participants peuvent s'y connecter à travers le réseau pour participer à un travail, une discussion, etc. Des figurines, plus ou moins réalistes, représentent les utilisateurs qui peuvent se déplacer dans la scène. Ils disposent d'outils communs tels que des tableaux, des planches à dessin, des agendas partagés, etc.

Les environnements « augmentés » de travail en groupe ont une symbolique très riche. L'image tridimensionnelle y apporte une force cognitive qui permet à chacun de comprendre immédiatement la nature d'une scène. En contrepartie, l'interaction est beaucoup plus difficile. Il est même quasiment impossible pour un débutant de naviguer convenablement dans ce type d'environnement.

Notre travail à Caen a consisté à construire des outils linguistiques pour intégrer des possibilités de conversation orale avec des assistants virtuels. Nous nous sommes concentrés sur un des points les plus délicats de l'interaction avec les mondes virtuels : la navigation. Nous avons construit un assistant – l'agent conversationnel *Ulysse* – doté de capacités de navigation et nous l'avons incorporé dans la représentation d'un utilisateur du monde. Avec ce modèle, l'utilisateur peut s'adresser à sa représentation et lui demander en son nom de se déplacer dans un environnement virtuel.

Le projet de recherche, qui s'est étendu sur quatre ans, a développé les modèles linguistiques nécessaires à la réalisation de l'agent et les a implantés. Ces modèles ont concerné le lexique, la syntaxe, la sémantique, la résolution de références, le raisonnement spatial et la planification. Ces domaines ont fait appel à beaucoup de concepts et de techniques abordés auparavant. Ils ont pu s'incorporer dans une entité unique permettant de naviguer à la voix dans un monde virtuel.

L'analyse lexicale a reposé sur l'examen de corpus de dialogue entre un utilisateur et l'agent du monde virtuel. Nous avons constitué ces corpus à Caen. De cette manière, on a pu se faire une meilleure idée de la phraséologie des utilisateurs dans ce type d'environnement : les locutions sont nombreuses, les phrases sont courtes, essentiellement affirmatives et centrées sur des verbes de mouvement. L'analyse lexicale d'*Ulysse* permet de détecter les locutions et nous avons défini des classes de verbes en fonction de types de mouvement.

L'analyse syntaxique de DictaMed était fondée sur des règles syntagmatiques. Le paradigme syntagmatique, pris dans sa totalité, est relativement rigide. Il suffit d'un mot qui ne soit pas dans le lexique ou d'une expression qui ne corresponde pas à la syntaxe pour que l'analyseur échoue. *Ulysse* a incorporé un analyseur fondé sur des techniques de chart incrémental.

Les charts permettent de gérer des hypothèses multiples et sont semblables aux ATMS. Le modèle du chart a été adapté à la parole pour sauter un mot d'hésitation ou pour accepter un mot inconnu. Il a été étendu pour réaliser des analyses partielles. Pour ceci, il a combiné un formalisme syntagmatique pour détecter les groupes nominaux et les groupes verbaux, et un formalisme de dépendance pour construire une représentation de la phrase complète, lorsque ceci est possible. Pour alléger les calculs inhérents à un chart, l'analyseur syntaxique a été divisé en niveaux dont l'architecture est celle d'un système multi-agents.

L'analyse sémantique consiste dans la construction d'une forme logique de l'énoncé. Pour ceci, *Ulysse* se fonde sur une analyse fonctionnelle des propositions. L'analyse fonctionnelle utilise des propriétés de composition et une grammaire de cas. Elle décompose les propositions formées autour d'un verbe complexe en une séquence équivalente exprimée en fonction de verbes plus simples. L'analyse sémantique peut réarranger la séquence de propositions en fonction des mots qui les relient : les connecteurs.

Les énoncés d'un utilisateur dans un univers virtuel comportent de multiples références à des objets. Ulysse résout ces références en utilisant la base de données d'objets qui modélise le monde virtuel. La résolution est multimodale. Elle prend en compte le champ de vision et les désignations à la souris. Elle transforme la liste des propositions en une séquence d'actions en se fondant sur un raisonnement géométrique impliquant pour chaque proposition, la préposition, le verbe et les objets.

À l'issue de l'étape du raisonnement géométrique, l'agent dispose d'une séquence de déplacements. Ils indiquent les coordonnées où l'agent doit se diriger et vers quel endroit regarder. « L'incarnation » d'Ulysse est constituée d'objets (tête, torse, bras, jambes, ...) articulés autour d'axes de rotations (épaules, bassin, cou, ...). Ulysse décompose son mouvement en actions élémentaires. Pour ceci, Il utilise une technique de planification linéaire. Grâce à elle, il peut générer une séquence de transformations des positions de tous les objets qui le compose. Ulysse se déplace en évitant les obstacles et il atteint sa position finale en pilotant une animation fluide et « humaine » de ses mouvements.

Nous avons conçu Ulysse dans le cadre du travail en groupe. Notre monde virtuel a d'abord représenté des bureaux. À la demande du groupe artistique *Das synthetische Mischgewebe*, nous avons ensuite construit un environnement virtuel fondé sur des images tridimensionnelles de cerveau. Nous avons adapté Ulysse à un nouveau vocabulaire et à un scénario de navigation. L'architecture d'Ulysse s'est révélée relativement génériques malgré le changement d'environnement. Un nouveau prototype en a résulté. Avec lui, un utilisateur peut se déplacer dans différentes zones d'un cerveau, les manipuler et questionner l'agent sur leurs fonctions. *Das synthetische Mischgewebe* en a créé un spectacle qu'il a présenté en France et à l'étranger.

1.4 Les étudiants que j'ai dirigés au cours du projet Ulysse et leur thème de travail

Dans le cadre du projet Ulysse, j'ai dirigé six étudiants de troisième cycle : deux thèses de l'Université de Caen, deux stages de diplôme d'études approfondies et deux *Diplomarbeit* allemands.

Dans l'ordre chronologique de réalisation, Korinna Diebel a construit le premier prototype de navigation (en allemand) pour son *Diplomarbeit* (1994). Korinna a utilisé le monde virtuel DIVE du SICS. Bien qu'elle n'ait disposé que d'une version incomplète de l'analyseur syntaxique, Korinna a pu faire fonctionner et montrer un premier système lors de la soutenance de son travail en 1994.

Pierre-Olivier El Guedj (1996) s'est occupé du module d'analyse syntaxique dans le cadre sa thèse. Il a intégré les deux paradigmes d'analyse : syntagmes et dépendances. Pierre-Olivier s'est aussi attaché à analyser un corpus de comptes rendus médicaux. Il a soutenu sa thèse en septembre 1996.

Christophe Godéreaux (1997) a rassemblé les corpus, réalisé le dépouillement lexical, écrit les grammaires et l'analyse sémantique. Christophe a intégré l'ensemble des modules dans DIVE et il a soutenu sa thèse en janvier 1997.

Frédéric Revolta (1995) et Olivier Bersot (1996) ont complété Ulysse dans le cadre de deux DEA. Frédéric a réalisé le résolveur de référence et le raisonnement géométrique en 1995. Olivier a réalisé le planificateur d'animations et les algorithmes de détection de collisions en 1996.

Ulysse utilise normalement plusieurs machines pour fonctionner : une ou deux stations de travail pour l'analyse linguistique et la synthèse d'images virtuelles du monde de bureau *Ithaque* et un PC pour la reconnaissance vocale. Matthias Ludwig (1996) a porté l'ensemble de l'agent sur un PC. Il l'a adapté à la navigation dans des images de cerveau et à leur manipulation. Matthias a réalisé ce travail dans le cadre d'un *Diplomarbeit* allemand.

1.5 Présentation du mémoire

Dans le mémoire qui suit, je décris les trois thèmes centraux des recherches que j'ai menées et que je poursuis. J'ai choisi de présenter dans chacun des thèmes, la réalisation la plus aboutie. Par exemple, l'agent conversationnel Ulysse, plutôt que le système de dialogue DictaMed, dans la mesure où le premier reprend et perfectionne les caractéristiques du second. Les thèmes sont les suivants :

- *l'architecture de systèmes de raisonnement*. L'organisation et la structure d'un système de raisonnement sont centrales pour son efficacité. Dans ce premier chapitre, je décris l'architecture d'un système multi-agents pour l'interprétation d'images réalisé au cours de ma thèse. Ce système a utilisé une architecture modulaire où chacun des modules avait un contour fixe. J'ai repris en partie les principes de ce premier système dans les domaines que j'ai abordés ensuite, d'abord dans l'analyse d'images puis dans l'analyse linguistique et les agents conversationnels. Je présente l'évolution de l'architecture des systèmes ultérieurs notamment pour prendre en compte la gestion d'hypothèses sous la forme de TMS, d'ATMS et de Charts.
- *l'analyse syntaxique*. L'analyse syntaxique est une étape essentielle pour construire la représentation d'une phrase et d'un texte. Dans ce second chapitre, je présente deux paradigmes de représentation de la syntaxe : les grammaires syntagmatiques et les grammaires de dépendance. Je décris comment nous avons combiné ces paradigmes au sein d'une architecture pour construire un analyseur. Je présente les résultats que nous avons obtenus sur un corpus médical. Le manque de corpus étiqueté et parenthésé en français rend une évaluation précise assez délicate. Je décris aussi dans ce chapitre l'analyseur syntaxique de l'anglais qui obtient les meilleurs résultats connus à ce jour. Cet analyseur présentant des similarités avec le nôtre.
- *les agents conversationnels*. L'analyse syntaxique de la langue ouvre la voie à sa compréhension. Pour l'instant, cette compréhension est limitée mais elle permet la création de nouvelles interfaces et de nouveaux outils. Dans ce dernier chapitre, je décris un agent conversationnel qui dispose de capacités syntaxique, sémantique et de raisonnement. Cet agent – Ulysse – se destine à aider un utilisateur à naviguer à la voix dans un monde virtuel. Il s'incorpore dans la représentation d'un utilisateur dans le monde. L'utilisateur peut alors commander à son agent de le déplacer – de

naviguer – en son nom. Je présente dans quel contexte Ulysse a été créé. Je détaille son architecture et les différents éléments qui le composent.

Enfin, en conclusion, je décris la problématique et les perspectives ouvertes par mon travail ainsi que le plan de mes recherches futures.

2.
Architecture de systèmes de
raisonnement :
application à l'interprétation
d'images

2.1 Présentation

Pendant ma thèse de doctorat, j'ai conçu un système pour identifier des protéines sur des images d'électrophorèse bidimensionnelle (Nugues 1993, Nugues & Haton 1989, Nugues 1989, Nugues et al. 1988a, Nugues et al. 1988b). L'interprétation de ces images nécessite de combiner des procédures de traitement du signal et de raisonnement. Pour ceci, le système a utilisé une architecture et des techniques à base de connaissances. Il a permis de reproduire les méthodes d'interprétation des biologistes et il a pu identifier sur une image des protéines connues; que celles-ci soient isolées ou à l'intérieur de constellations. Cette architecture a servi de référence pour celles que j'ai adoptées par la suite au cours d'autres projets.

L'électrophorèse est une technique d'analyse biologique. Elle permet de séparer les protéines d'un tissu suivant leurs caractéristiques physico-chimiques. L'électrophorèse bidimensionnelle (O'Farrell 1975) exploite deux processus de séparation dans des directions orthogonales. Elle sépare les protéines suivant leur point isoélectrique (pI) selon l'axe des x et leur masse moléculaire selon l'axe des y (Figure 1) :

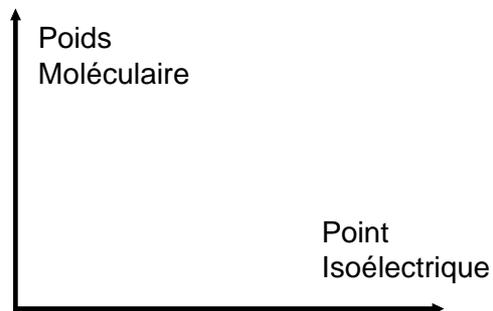


Figure 1 Les axes électrophorétiques.

Un gel d'électrophorèse est transparent à l'origine et des colorants sont nécessaires pour révéler les protéines. En 1987-1989, c'était principalement le bleu de Coomassie, l'argent et les éléments radioactifs. Après la révélation, le gel prend l'aspect d'un buvard taché (Figure 2) dont les taches représentent les protéines. Selon les tissus, les gels peuvent être plus ou moins lisibles. Certaines taches sont isolées et clairement visibles, d'autres sont plus difficiles à identifier.

Figure 2 Un gel de plasma humain.

Les distributions statistiques des protéines autour de leur point caractéristique peuvent se modéliser comme des distributions gaussiennes le long de l'axe des x et de l'axe des y . L'Équation 1 donne le modèle de base d'une protéine sur une image en niveau de gris. Dans cette équation, (x_0, y_0) est le centre de la tache et (σ_x, σ_y) son écart-type.

$$I(x,y) = A \exp \left(-\frac{(x-x_0)^2}{2\sigma_x^2} - \frac{(y-y_0)^2}{2\sigma_y^2} \right)$$

Équation 1 Les niveaux de gris.

Le système d'analyse et d'interprétation a adopté une architecture de type multi-agents. Il comprenait des procédures de traitement d'images ainsi que des méthodes de raisonnement ascendant et descendant. Le traitement d'images avait pour but de réaliser l'extraction des paramètres des taches sur l'image tandis que le processus de raisonnement devait appairer ces paramètres aux modèles géométriques des

configurations de protéines. Après ce cycle, l'architecture permettait de retourner sur l'image pour retrouver d'éventuelles protéines manquantes.

2.2 Identifier des protéines sur des gels

2.2.1 L'intérêt et l'utilisation de l'électrophorèse bidimensionnelle

La structure des cellules de tous les êtres vivants est déterminée par son information génétique. Cette information, qui est portée par l'ADN, s'exprime dans chaque cellule grâce à un code qui détermine une séquence d'acides aminés. Les acides aminés forment des polypeptides et des protéines. Les protéines d'un tissu définissent son type, c'est-à-dire sa structure et son activité. On estime le nombre de protéines à 100.000 chez l'homme et un tissu peut en contenir plusieurs milliers.

L'électrophorèse bidimensionnelle produit une signature de la composition protéique d'un tissu. Elle révèle notamment les variations de composition d'un même tissu ou d'un liquide corporel entre différents individus. On peut relier certaines de ces variations à un état clinique ou pathologique. De plus, l'électrophorèse permet d'étudier la fonction d'une protéine ou d'un ensemble de protéines d'un tissu en faisant varier certains des paramètres de l'environnement, par exemple en le soumettant à des produits chimiques ou à des contraintes physiques.

L'électrophorèse bidimensionnelle a grandement étendu le domaine de l'analyse des protéines. Comme conséquence, les gels des tissus peuvent révéler un certain nombre de taches non identifiées. Le dépouillement d'un ensemble de gels se fonde en général sur la comparaison de gels de tissus témoins avec ceux d'autres tissus. La comparaison s'effectue en reliant les taches d'emplacements et d'aspects semblables le long d'une séquence de gels (Figure 3).

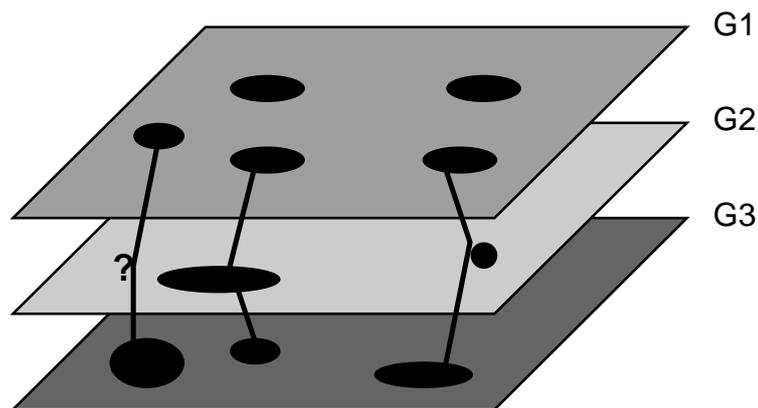


Figure 3 Le recalage de gels.

2.2.2 La comparaison classique et ses problèmes

Plusieurs méthodes de comparaison de gels ont été proposées. Elles se fondent toutes sur l'appariement manuel d'un certain nombre de points d'ancrage. On doit distribuer ces points d'ancrage relativement régulièrement sur la surface du gel et on doit les identifier sans ambiguïté sur l'ensemble des gels. On choisit les points d'ancrage, en général, pour leur grande taille.

L'algorithme de comparaison décrit automatiquement l'environnement des points d'ancrage, c'est-à-dire les taches autour d'eux, par des relations géométriques. Les environnements respectifs sont comparés deux à deux. Les taches entourant les points d'ancrage qui se trouvent à l'intérieur de la description sont appariées d'un gel à l'autre suivant des critères de similitude géométrique. L'algorithme itère ce processus sur les taches les plus récemment appariées. Chaque appariement est éventuellement suivi d'une transformation géométrique sur l'environnement en jeu. La transformation géométrique, par exemple une déformation bilinéaire, réalise un recalage élastique en étirant ou en comprimant des régions de l'image.

On observe des variations dans l'emplacement des protéines entre les gels et on doit définir différents seuils pour déterminer si l'appariement entre deux taches doit se faire ou non. Ces seuils doivent tenir compte de la reproductibilité du procédé électrophorétique. Ils sont difficiles à régler car ils doivent gérer le compromis suivant :

- être strict, avec le risque que des taches homologues ne soient pas appariées;
- être lâche, avec le risque d'un mauvais appariement.

L'algorithme d'appariement procède à une analyse globale. Il peut produire pas mal d'erreurs, principalement parce qu'aucun modèle *a priori* n'est utilisé sur les configurations possibles des protéines. Ces mauvais appariements sont difficiles à contrôler et sans connaissance *a priori* des modèles géométriques, on ne peut faire que des statistiques sur la validité de l'appariement. Elles conduisent à une grande incertitude dès qu'un nombre de plus d'une dizaine de gels est en jeu. Par exemple, un taux de mauvais appariement de 1% entre deux gels conduit à un taux de réussite de $(1 - 0,1)^{40} \approx 0,67$ quand on réalise 40 appariements correspondant à une série de 40 gels. Ce taux n'est pas acceptable pour une exploitation de routine.

2.2.3 L'identification en utilisant des connaissances préalables

Pour certains tissus, on dispose d'une connaissance sur l'aspect du gel correspondant. Des biologistes sont capables de reconnaître des taches isolées ou bien des groupes de plusieurs taches voisines (des constellations). On peut parfois identifier certaines protéines par leur nom, sans avoir à les comparer avec d'autres gels. Certaines constellations présentent des différences entre les individus et on peut parfois relier des variations de configurations avec des maladies ou bien l'état particulier d'un patient.

L'étendue de cette connaissance humaine, qu'on obtient par des entretiens avec un expert, peut concerner une partie importante des taches d'un gel, surtout s'il s'agit

d'un tissu ou d'un liquide corporel très étudié comme le plasma. Le processus de raisonnement de l'expert quand il identifie les taches met en jeu différentes étapes :

1. *une focalisation sur une région particulière* du gel où il doit trouver les taches en question; par exemple le coin bas à gauche du gel;
2. *la reconnaissance d'une forme générale* de l'environnement; par exemple, trois alignements parallèles ou deux taches importantes en diagonale.
3. *une identification* des taches dans cet environnement.

Avec un système à base de connaissances, l'ordinateur imite cette méthode d'identification d'une tache ou d'un groupe. La connaissance experte doit être alors acquise et représentée dans la machine puis reproduite automatiquement suivant les trois points précédents.

Cette connaissance comprend de l'expertise et des procédures à des niveaux conceptuels différents :

- le niveau *perceptif*, tel que la détection des taches qui doit être implanté par des procédures de traitement d'images;
- le niveau *géométrique*, tel que l'appariement de constellations à des modèles, en tolérant des taches manquantes et des variations locales. Certaines constellations montrent un polymorphisme, c'est-à-dire qu'il y a différents modèles d'emplacements de taches entre les individus. La méthode que nous avons utilisée demande que l'expert détaille toutes les formes connues d'une constellation et ses variantes. Ce niveau peut s'implanter en utilisant des règles géométriques.
- Le niveau *interprétatif*, tel que la vérification de la cohérence générale entre les protéines identifiées dans les différentes régions. L'expert peut éventuellement proposer un diagnostic sur l'état du patient ou du tissu. Ce niveau peut s'implanter par des règles de déduction.

Par rapport à l'appariement semi-automatique, lorsqu'une expertise est disponible, la méthode utilisant des connaissances externes possède les avantages suivants :

- on peut identifier les taches ou les groupes de taches sur un seul gel et on n'a plus besoin de procéder à une comparaison dans une série.
- l'analyse est locale : chaque région est appariée à un modèle géométrique déterministe.
- Tous les modèles possibles de constellations (les modèles géométriques) sont détaillés et les variations d'emplacement des taches sont clairement permises et définies.
- Enfin, on peut relier certaines constellations avec des interprétations biologiques ou médicales.

De plus, quand l'expertise disponible à un moment donné ne couvre pas la totalité des taches d'un gel, cette méthode peut s'utiliser comme une première identification partielle, couvrant les taches connues, et suivie d'un appariement classique sur le reste du gel.

2.3 Une architecture pour identifier des protéines

2.3.1 Les principes

L'architecture implantée est fondée sur un modèle multi-agents (Engelmore & Morgan 1988, Haton et al. 1991). Elle prend la forme d'un système de règles de production hiérarchique divisé en *Sources de connaissances* (SC). Cette architecture intègre des procédures et des règles pour réaliser le traitement d'images et l'identification des protéines sur les gels d'électrophorèses bidimensionnelles.

Les SC se composent de spécialistes qui prennent la forme d'un ensemble de règles. Les règles peuvent comprendre des procédures dans leur partie action. Les spécialistes sont groupés dans des SC en fonction de la similitude de leur tâche : traitement d'images, analyseurs de formes ou systèmes de déduction. Une source de connaissances particulière – le pilote – réalise la gestion des autres et les active. À l'intérieur des sources de connaissances, les spécialistes sont déclenchés en fonction d'un agenda.

Le système n'opère pas sur tout le gel en entier. Il se concentre sur une partie par un focaliseur d'attention. La focalisation de l'attention se centre d'abord sur les régions les plus simples du gel, où les taches sont faciles à identifier ou clairsemées. Les régions de focalisation sont choisies par l'expert comme point de départ de son raisonnement. Une fois que les protéines sont identifiées dans le cadre de focalisation, le recalage des protéines du cadre dans des coordonnées normalisées est possible. Les régions de focalisation sont ensuite déplacées vers des zones de plus en plus complexes.

Un raisonnement ascendant exploite les sources de connaissances dans un premier temps. Elles déterminent l'emplacement des taches en utilisant des procédures de traitement d'images et puis tentent d'apparier ces taches aux modèles d'emplacement de protéines ou de formes de constellation. Les taches de l'image sont appariées aux modèles en utilisant un retour en arrière si nécessaire.

Comme les images sont parfois dégradées, l'appariement est graduel. On apparie d'abord les pics les plus grands et les plus constants dans le modèle géométrique avec les pics de l'image, suivis des pics plus petits et plus instables.

Dans certains cas, l'appariement est partiel; par exemple, quand des petits pics disparaissent, recouverts par d'autres proches et plus grands ou lorsqu'ils sont trop petits pour être détectés avec un niveau de confiance suffisant. L'emplacement dans le modèle reste alors non instancié.

Un raisonnement descendant est ensuite mis en œuvre. Il vérifie la validité des constellations et leur cohérence générale. Dans certains cas, le pilote peut l'activer pour retrouver une tache manquante sur l'image. Par exemple, si des emplacements de taches ne sont pas proprement remplis, on tentera de les prédire. Ceci déclenchera alors une procédure de traitement d'images pour détecter un maximum autour de l'emplacement prédit avec un seuil de détection plus faible.

2.3.2 Les sources de connaissances

Le système est constitué de quatre sources de connaissances génériques autour du pilote (Figure 4). Les SCs contiennent des spécialistes (des ensembles de règles) qui peuvent réaliser le même type d'action ou tirer le même type de conclusion. Les sources de connaissances sont clonées pour chacune des constellations à retrouver sur l'image. C'est-à-dire que chacune des constellations, comportant un ou plusieurs modèles d'emplacement de taches, est traité par quatre sources de connaissances particulières dérivées des quatre génériques.

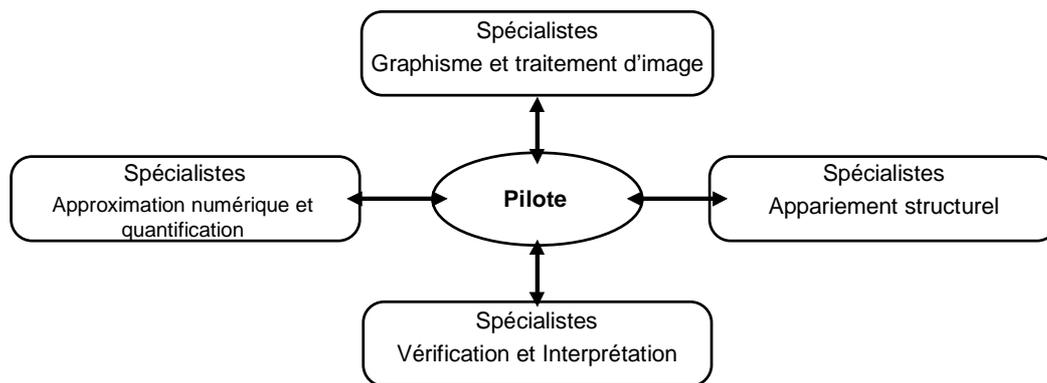


Figure 4 L'architecture du système.

Les sources de connaissances ont les fonctions suivantes :

- graphisme, accès à l'image, focalisation, traitement d'image et détection de pics (GTI);
- approximation numérique et quantification (ANQ);
- appariement structurel complet ou partiel (AS);
- vérification et interprétation (VI).

Le système traite les constellations l'une après l'autre dans un ordre déterminé et pour chacune des constellations, le pilote commande la séquence d'activation des sources de connaissances correspondantes. Il élit une SC à la fois et attend qu'elle se termine. Le pilote détermine la séquence des sources de connaissances suivant :

- ses *buts*, se focaliser sur les régions des constellations dans un ordre de complexité croissante et pour chaque constellation, effectuer la séquence de traitement suivant : traitement d'image, appariement, interprétation et quantification (Figure 5, flèches du haut).
- certaines *données* ou événements en provenance des sources de connaissances correspondant à un échec d'appariement, une incohérence ou une tâche manquante, qui modifient la séquence précédente (Figure 5, flèches du bas).

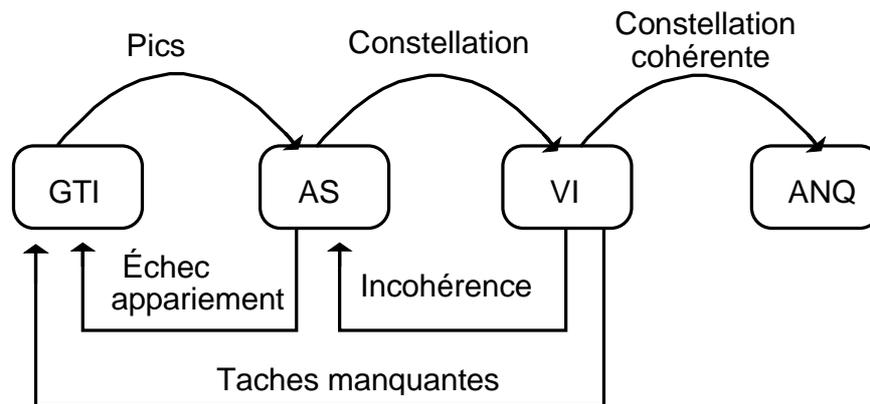


Figure 5 Diagramme d'activation des sources de connaissances.

Suivant le type des événements, on entreprendra les actions suivantes :

1. Un échec dans l'appariement (Source de connaissances AS) conduira à un élargissement de la région de focalisation avec un abaissement du seuil de détection.
2. Une incohérence détectée dans la source de connaissances de vérification et d'interprétation conduira à la destruction de l'instance du modèle et à l'activation de la source de connaissances d'appariement structurel.
3. Un point manquant par rapport au modèle de constellation pourra voir son emplacement prédit par la source de connaissances de vérification et d'interprétation. Dans ce cas, la source de connaissances de traitement d'image sera activée pour tenter de le retrouver dans la région prédite.

2.3.3 Objets et communications

Les sources de connaissances manipulent des objets tels que des régions d'image, des pics, des pics identifiés, des constellations, etc. Ces objets appartiennent à des classes. Les objets de la classe « pic » comprennent, par exemple, une amplitude, une abscisse, une ordonnée qui peuvent être partiellement ou complètement intanciées.

Les règles des spécialistes comportent des variables dans leur partie condition et action. Pour déclencher une règle, un objet doit s'unifier aux variables de la partie condition en respectant certaines contraintes. Dans la partie action, les règles peuvent créer, modifier ou détruire des objets. De plus, les règles peuvent aussi inclure des procédures qui entrent dans leur domaine de compétence.

La communication entre les sources de connaissances et le pilote s'opère par des messages qui se représentent comme des classes d'objets. Chaque SC communique par une classe particulière de message; c'est-à-dire qu'il y a une classe de messages pour les communications entre le pilote et la SC Graphisme et traitement d'image (GTI), entre GTI et le pilote, entre le pilote et l'Appariement structurel, etc. Les messages du pilote peuvent notamment contenir un agenda pour la source de connaissances de réception et certains paramètres. Les messages des sources de

connaissances contiennent des descriptions du résultat qu'elles ont obtenu. La structure de ces messages s'est inspirée de celle de Gong & Haton (1988).

2.3.4 Agendas

La stratégie d'une SC particulière est donnée par son agenda qui lui spécifie la suite de spécialistes applicables. Un élément de l'agenda peut donc définir une ou plusieurs règles qui seront inférées en chaînage avant. L'agenda est dynamique; il est créé, modifié ou bien détruit par la partie action d'une règle (Figure 6) :

1. Chaque source de connaissances contient un agenda par défaut qui correspond à l'exécution normale.
2. Le pilote peut créer un agenda pour la SC à activer; cet agenda est déterminé suivant les données courantes connues du pilote.
3. La source de connaissances active peut modifier ou changer son agenda suivant les faits créés par une procédure de traitement d'image ou bien déduit par une règle. Les faits doivent alors être locaux à la SC.

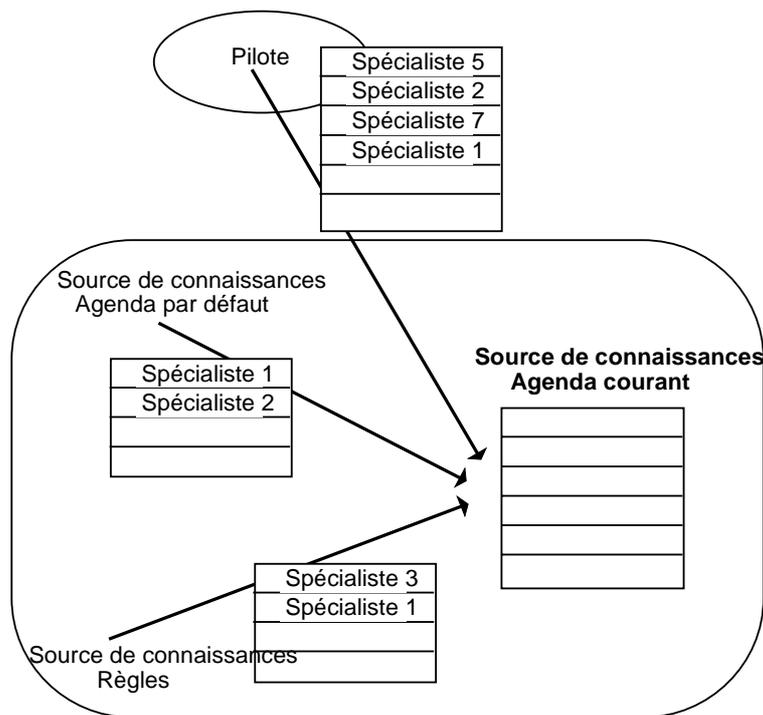


Figure 6 Les agendas.

L'utilisation des agendas permet de restreindre l'étendue du chaînage en avant ainsi qu'une définition souple du pilotage de chaque source de connaissances.

2.4 Traitement d'images

La source de connaissances Graphisme et traitement d'image (GTI) contient un ensemble de spécialistes gérant les opérateurs de traitement d'image. À la première activation, l'agenda par défaut détermine les opérations. La source de connaissances GTI se focalise sur une région de taille rectangulaire où les taches de la constellation recherchée sont les plus probables. L'emplacement des taches est déterminé en utilisant des règles d'expertise qui gèrent des coordonnées absolues ou relatives à des taches identifiées précédemment sur le gel. Elles sélectionnent, par exemple, le quart inférieur gauche du gel ou une zone bien définie au-dessus d'une tache particulière.

À l'intérieur de la région sélectionnée, la SC peut effectuer un lissage pour éliminer le bruit de haute fréquence. Elle détecte ensuite les pics correspondant à des taches isolées ou connexes au-dessus d'un seuil déterminé.

La source de connaissances GTI peut ensuite être activée par le pilote par un retour en arrière si l'identification ultérieure a échoué ou sur une analyse descendante si une tache manquait. Dans le premier cas, la SC élargirait la région d'analyse et abaisserait le seuil de détection. Dans le second cas, elle chercherait à détecter une tache dans une sous région particulière.

Le message du pilote contient certains arguments nécessaires pour exploiter la source de connaissances GTI : l'agenda, le seuil, le coefficient d'agrandissement, les coordonnées du cadre de focalisation. Seuls les arguments utiles sont passés. Par exemple, les coordonnées du cadre de focalisation ne sont transmises que sur une tentative, lors d'une analyse descendante, de retrouver une tache spécifique dans une région déterminée.

L'agenda de la source de connaissances contient la liste des différents spécialistes dans leur ordre d'activation. Les spécialistes principaux sont les suivants :

- *l'initialisation* qui affecte certains paramètres, notamment le seuil de détection des pics à moins qu'il ne soit fourni dans le message. L'initialisation active ensuite l'agenda par défaut ou bien le remplace par l'agenda contenu dans le message lorsque le pilote en transmet un.
- La focalisation avec trois variantes possibles :
 1. la *focalisation automatique* qui découpe la région la plus probable d'une constellation en utilisant des coordonnées absolues ou relatives
 2. *l'agrandissement de la région* qui découpe la région à nouveau en l'agrandissant d'un coefficient déterminé
 3. la *focalisation paramétrée* qui découpe la région en utilisant les coordonnées de cadre de focalisation provenant du message du pilote.
- Le *lissage* qui élimine le bruit de haute fréquence;
- la *détection des pics* qui détecte les extrema dans le cadre de focalisation au-dessus d'un seuil déterminé. Les instances de pics sont alors créées avec leurs amplitude, abscisse et ordonnée.
- la *terminaison* qui rend les commandes au pilote. Cependant en cas de résultats anormaux, notamment si aucune tache n'est trouvée, le spécialiste *terminaison* peut

de nouveau activer la source de connaissances GTI en changeant d'agenda. Le spécialiste *terminaison* utilisera alors un seuil de détection plus faible et un coefficient d'agrandissement. Le nouvel agenda comprendra *l'agrandissement de la région, détection des pics, et terminaison* lui-même.

2.5 Identification des protéines

L'identification des protéines, ou des constellations de protéines, est réalisé par la SC d'appariement structurel. Elle utilise les pics détectés dans la région de focalisation précédente et en filtre une configuration potentielle.

Quand la constellation est unique, c'est-à-dire qu'il n'y a pas de variante, cette configuration doit correspondre à un appariement partiel ou total au modèle géométrique. L'appariement est partiel quand il manque une ou plusieurs taches sur l'image. Plusieurs modèles correspondant à la configuration complète ou bien partielle sont utilisés pour réaliser l'appariement.

On recherche le modèle en utilisant le retour en arrière en cas d'échec. Initialement, on recherche le modèle le plus complet sur l'image, puis des modèles partiels, jusqu'au modèle minimal. Le modèle minimal d'une configuration partielle comprend les pics qui sont toujours visibles sur le gel quand la constellation correspondante est présente.

L'appariement entre les taches et les modèles se fait en utilisant des conjonctions de conditions dans les prémisses des règles. Ces conjonctions contiennent des variables à instancier – à relier à des valeurs extraites du gel – et elles expriment des contraintes géométriques sur l'emplacement des taches et sur leur amplitude. L'appariement est graduel : les taches les plus hautes sont traitées d'abord. Viennent ensuite les autres pics qui sont ajoutés au modèle.

Les règles qui expriment le modèle complet décrivent géométriquement la présence des taches en relation avec leur environnement. Les règles d'appariement partiel utilisent les mêmes contraintes mais spécifient de plus l'absence d'une ou plusieurs taches en comparaison avec le modèle complet.

Un exemple simplifié de règles d'appariement correspondant à la détection de S1 et de S3 de la Figure 7 est donné sur le Tableau 1. Les conditions **C1**, **C2**, **C3** et **C4** filtrent les deux taches : **S1** et **S3**, en diagonale avec des tolérances. Les variables à instancier commencent avec un « ? » (?x par exemple). **C1** et **C2** appartiennent dans la classe « pic » la tache ?p1 avec l'amplitude ?a1, l'abscisse ?x1, et l'ordonnée ?y1 de telle sorte qu'il n'y ait pas de pic plus grand dans l'intervalle rectangulaire $[-valx1, +valx1] \times [-valy1, +valy1]$ centré autour de lui. **C3** et **C4** appartiennent la plus grande tache ?p3 avec l'amplitude ?a3, l'abscisse ?x3 et l'ordonnée ?y3 située dans l'intervalle rectangulaire $[valx2, valx3] \times [valy2, valy3]$ relativement à ?p1.

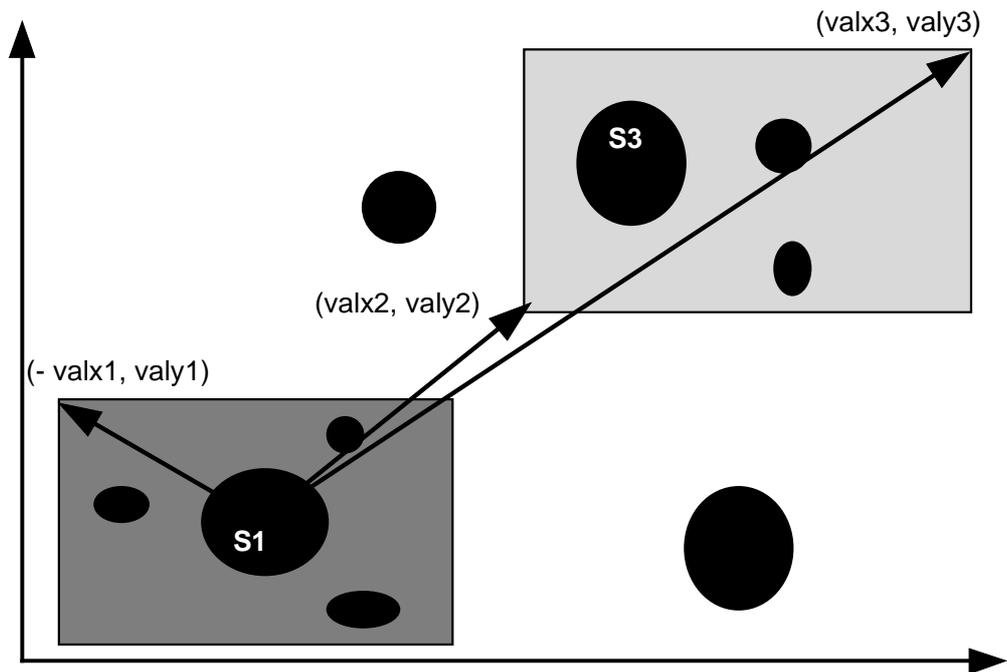


Figure 7 Un exemple de configuration de taches.

```

C1
(class pic ?p1 (amplitude ?a1) and (abscisse ?x1) and (ordonnée ?y1))
C2
(with no
(class pic ?p2 (amplitude ?a2 (constraint (> ?a1)))
and (abscisse ?x2 (constraint (> (?x1 - valx1)) and (< (?x1 + valx1))))
and (ordonnée ?y2 (constraint (> (?y1 - valy1)) and (< (?y1 + valy1))))))
C3
(class pic?p3 (amplitude ?a3)
and (abscisse ?x3 (constraint (> (?x1 + valx2)) and (< (?x1 + valx3))))
and (ordonnée ?y2 (constraint (> (?y1 + valy2)) and (< (?y1 + valy3))))))
C4
(with no
(class pic?p4 (amplitude ?a4 (constraint (> ?a3)))
and (abscisse ?x4 (constraint (> (?x1 + valx2)) and (< (?x1 + valx3))))
and (ordonnée ?y4 (constraint (> (?y1 + valy2)) and (< (?y1 + valy3))))))

```

Tableau 1 Des règles d'appariement.

2.6 Vérification et interprétation

La configuration trouvée par la source de connaissances d'Appariement structurel est évaluée par celle de Vérification et d'interprétation. Cette source de connaissances vérifie la cohérence de la constellation courante avec le reste de l'environnement (les autres extrema de la région de focalisation) et les constellations précédemment identifiées.

La SC de vérification et d'interprétation peut rejeter la variante trouvée. Elle détruira alors l'instance et reviendra en arrière et rendant le contrôle au pilote. Le pilote construira alors un nouvel agenda en considérant les variantes suivantes de la constellation et il activera la SC d'appariement structurel. Si aucun modèle n'est trouvé, un autre retour en arrière sera tenté sur la SC GTI pour agrandir la région de focalisation.

Si la constellation est évaluée avec succès, la SC de vérification et d'interprétation déterminera les taches manquantes et pourra prédire leurs emplacements approximatifs. Une analyse descendante tentera alors de les retrouver grâce à la SC GTI – avec un seuil de détection plus bas – dans des zones qui entourent ces emplacements.

Finalement, on pourra relier certaines configurations de protéines à un diagnostic médical. La SC de vérification et d'interprétation permet d'ajouter des règles de diagnostic qui peuvent exister.

2.7 Application aux apolipoprotéines du plasma

Nous avons réalisé un prototype du système pour identifier les apolipoprotéines du plasma. Les apolipoprotéines jouent un rôle dans le métabolisme des lipides et des variantes particulières sont impliquées dans les maladies cardio-vasculaires. Nous avons considéré les apolipoprotéines A-I, A-II, C-2 et C-3 pour lesquelles nous avons déterminé les régions de focalisation, les méthodes d'accès, les emplacements des taches avec leurs variantes. Ce travail s'est fait en collaboration avec le Centre de médecine préventive de Vandœuvre-lès-Nancy. L'expertise a été acquise durant des entretiens avec une pharmacienne habituée à l'interprétation visuelle des gels d'électrophorèse.

Les biologistes considèrent d'abord la région de la préalbumine, ainsi que les chaînes d'haptoglobines α -1 et α -2 pour prendre des points d'ancrage. Le système reproduit ce processus avec un premier groupe de sources de connaissances : les SC haptoglobine. La SC GTI haptoglobine focalise son attention sur cette région qui couvre à peu près le quart inférieur gauche du gel. Elle détecte l'ensemble des pics compris dans ce cadre.

Le contrôle est passé à la source de connaissances d'appariement structurel et les modèles sont instanciés graduellement et éventuellement partiellement, en utilisant les pics trouvés jusqu'à obtenir un modèle. La SC d'appariement structurel des haptoglobines contient plusieurs modèles notamment pour tenir compte de leur polymorphisme. Par exemple, sur la Figure 2, l'haptoglobine α -2 n'est composé que d'une tache clairement visible alors chez d'autres sujets il y a un alignement de plusieurs taches. Les taches sont identifiées et l'une d'elles est d'une importance particulière : la grande tache de la chaîne de préalbumine. Son abscisse sert de point de référence pour l'identification des constellations suivantes. En cas d'échec, le contrôle est rendu à la SC GTI qui agrandira le cadre de focalisation.

La SC de Vérification et d'interprétation de l'haptoglobine vérifie la cohérence des taches appariées avec celles qui restent. En cas d'incohérence, on opère un retour en arrière sur la SC d'appariement structurel. Pour certaines instanciations partielles, par

exemple si la tache gauche de la préalbumine manque, la SC GTI sera activée pour les retrouver. Finalement, la SC ANQ quantifie les taches.

Quand le cycle est terminé, d'autres constellations sont traitées en utilisant le même principe :

1. vers le bas, les apolipoprotéines de poids faibles : apocC et apoA-II, si elles sont présentes en utilisant leurs polymorphismes les plus courants.
2. vers le haut, la constellation des apoA-I en commençant avec la proapolipoprotéine A-I dont l'abscisse est approximativement égale à celle de la grande tache de la préalbumine.

La Figure 8 donne les étapes simplifiées de l'identification de la chaîne d'apoA-I :

1. une région est extraite contenant la préalbumine, les chaînes d'haptoglobines α -1 et α -2 où les procédures d'images identifient les extrema (SC GTI haptoglobine);
2. Les règles d'identification structurelle identifient les alignements de taches qui correspondent à la préalbumine et aux chaînes d'haptoglobines α -1 et α -2 (SC AS haptoglobine);
3. La grande tache de la chaîne de préalbumine est identifiée. Son pI permet de déterminer un nouveau cadre de focalisation contenant la tache de la proapolipoprotéine A-I et la chaîne d'apolipoprotéine A-I (SC GTI ApoAI);
4. La tache de proapolipoprotéine A-I est déterminée grâce au pI de la préalbumine et simultanément avec l'existence de la chaîne d'apoA-I juste en dessous (SC AS ApoAI).

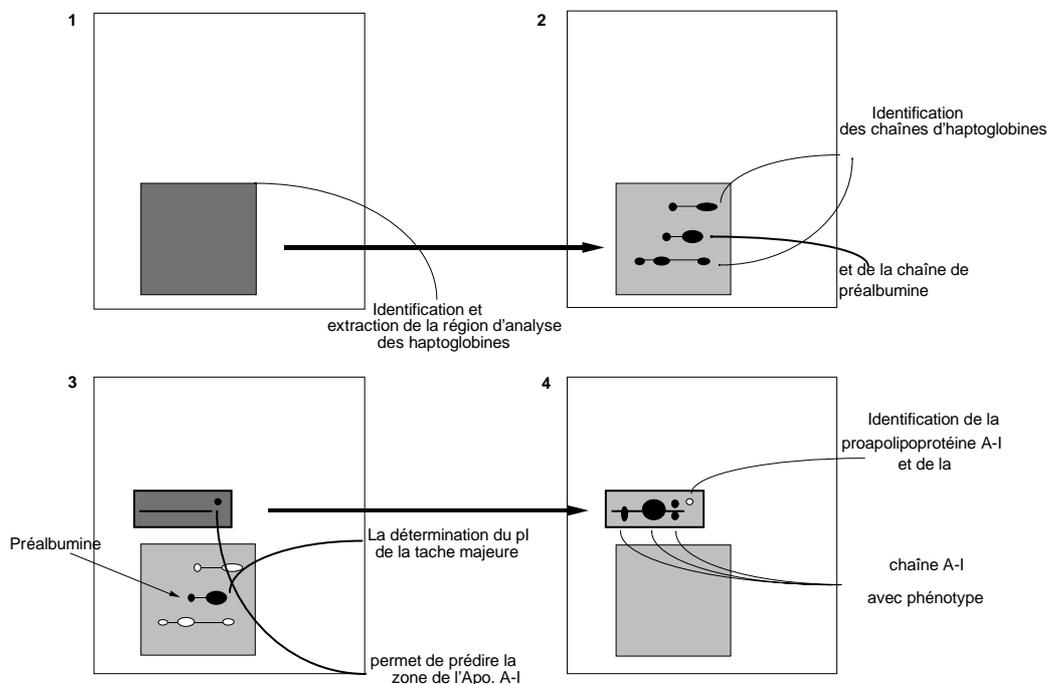


Figure 8 L'identification simplifiée de l'apolipoprotéine A-I.

J'ai pu évaluer le prototype sur cinq gels de qualité médiocre fournis par le Centre de médecine préventive. Ce choix a été fait à dessein, car il représente une sorte de pire cas d'analyse quotidienne clinique. Les chaînes de préalbumine et d'haptoglobine α -1 et α -2 tout comme les apolipoprotéines A-I, A-II, C-2 et C-3 ont été identifiées avec succès à l'exception de l'Apo A-I d'un gel où le système a étiqueté une chaîne en dessous.

2.8 Conclusion

2.8.1 Perspectives pour l'électrophorèse

Au cours de ma thèse, j'ai conçu et implanté un prototype pour interpréter des électrophorèses bidimensionnelles. Son architecture s'est fondée sur des techniques multi-agents. Le système a permis la reproduction d'une connaissance humaine obtenue auprès de biologistes pour identifier des protéines sur un gel. Il permet aussi de relier certaines configurations de protéines à des risques ou à une pathologie quand la connaissance sur le sujet est disponible.

Une des limites que j'avais décrites à la fin de ma thèse se trouvait dans les lacunes potentielles de la connaissance : des configurations qui n'auraient pas été détaillées par l'expert. On peut se trouver en présence, par exemple, de variantes rares ou inconnues. Le modèle correspondant ne sera alors pas dans la base de connaissances. Dans le meilleur des cas, la forme de cette configuration ne s'appariera à aucun modèle. Le système échouera à l'identifier et la rejettera comme inconnue. Cependant, les modèles géométriques contiennent des tolérances qui peuvent toujours faire tomber la nouvelle constellation dans une mauvaise description.

En 1988, l'électrophorèse bidimensionnelle était assez imprécise et d'une exploitation plutôt difficile. Depuis sa reproductibilité est devenue bien meilleure et des kits maniables sont maintenant disponibles. Après une période où elle a suscité un intérêt moins grand, sa diffusion s'étend de nouveau. Ceci est en grande partie dû au séquençage en masse de génomes. Des séquences génétiques complètes sont disponibles pour certains organismes. En parallèle se développent des cartes électrophorétique de plus en plus précises comme celle de l'E. Coli (VanBogelen et al. 1996). L'intérêt des biologistes se porte de plus en plus sur le déchiffrement des fonctions des gènes : comment et quand ils s'expriment. Pour ceci, l'électrophorèse bidimensionnelle est une technique de choix.

La meilleure reproductibilité du procédé électrophorétique rend méthode d'interprétation fondée sur l'expertise plus fiable. Actuellement, on compare encore les gels d'une même série. La technique fondée sur des connaissances préalables permet un étiquetage partiel des protéines d'un gel et leur identification. De plus la disponibilité de grandes bases de données sur les protéines ouvre la voie à un étiquetage complet. J'avais exposé cette potentialité à la fin de ma thèse (Nugues 1988b). Il est vraisemblable que bientôt, il sera possible de comparer un gel à une base distante pour évaluer les différences d'expression des protéines et les comparer à des résultats obtenus ailleurs et à un autre moment. J'ai entamé une coopération à

ce sujet – encore embryonnaire – avec Francis Biville de l’Institut Pasteur sur des protéines de l’E. Coli.

2.8.2 L’évolution de l’architecture

2.8.2.1 DictaMed

Après avoir obtenu un poste de Maître de conférences à Caen, je me suis consacré à l’informatique linguistique et dans un premier temps à la conception du système de dialogue oral DictaMed (Nugues et al. 1992; Nugues et al. 1993(a); Nugues et al. 1993(b); Cazenave et al. 1994).

DictaMed a utilisé un dispositif de reconnaissance vocale de mots isolés du commerce ainsi qu’un circuit de synthèse de parole. DictaMed acceptait les énoncés de l’utilisateur et il en réalisait une analyse syntaxique. Il construisait simultanément une représentation sémantique des phrases. DictaMed avait la capacité d’interagir oralement avec l’utilisateur. L’analyseur syntaxique, notamment, acceptait les formes interrogatives et, grâce à l’analyseur sémantique, DictaMed pouvait répondre à certaines questions concernant des entités précédemment décrites.

DictaMed a divisé et organisé les connaissances de syntaxe, de sémantique et de dialogue dans une structure modulaire. Son architecture s’est appuyée sur un système multi-agents. Elle a largement hérité de la structure du système d’interprétation d’électrophorèses (Figure 9).

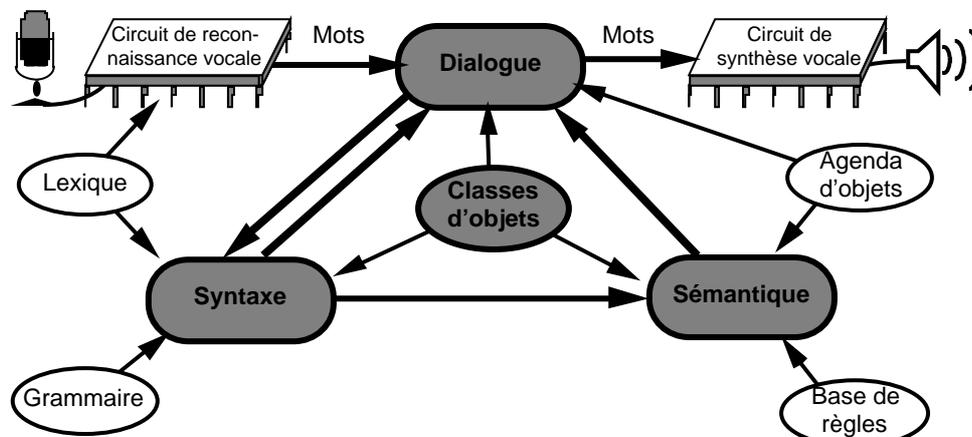


Figure 9 L’architecture du système de dialogue oral DictaMed.

La méthode de recherche de solutions pour identifier les constellations de protéines aussi bien que celle de l’analyse linguistique de DictaMed ont reposé sur le retour en arrière en cas d’échec.

Une autre gestion du raisonnement est possible en émettant et en explorant des hypothèses en parallèle. Un des algorithmes de raisonnement hypothétique est celui de l’ATMS – *Assumption-based Truth Maintenance System* – (de Klerer 1986). La recherche de solutions se fait en considérant plusieurs hypothèses à la fois au lieu d’une seule. On développe les hypothèses par un raisonnement en utilisant des règles. Il en résulte à chaque étape et pour chaque hypothèse un ensemble de faits

qu'on peut fusionner deux à deux. On peut par ailleurs poser des contraintes générales de contradiction sur les faits :

Si Fait A et Fait B alors faux.

Si au cours du développement d'une des hypothèses ou d'une fusion Fait A et Fait B apparaissent ensemble, alors l'hypothèse correspondante sera éliminée.

2.8.2.2 Le recalage d'images radar

À la suite de l'interprétation d'électrophorèses, j'ai construit un prototype pour recalcr des images de radar sur une carte géographique. L'organisation générale du système a été la même que pour l'interprétation des électrophorèses et elle a associé des opérateurs de traitement d'image et des méthodes de raisonnement. Dans le cas d'une image radar cependant, les objets pertinents ne sont plus les taches mais les routes. Le recalage nécessite de comparer les paramètres des routes d'une carte à celle des routes d'une image. Les opérateurs de traitement ont donc comporté des détecteurs de lignes droites fondés sur la transformation de Hough.

Le processus de raisonnement a utilisé un ATMS. Il générait les hypothèses d'appariement entre les segments rectilignes de route de la carte et ceux obtenus par la transformation de Hough de l'image radar. Le développement du raisonnement combinait les routes incrémentalement. S'il se produisait une incohérence d'orientation à un moment donné, l'hypothèse était écartée définitivement, sinon elle restait valide. Sur les quatre images de test, trois ont pu être recalées avec succès (Amat et al. 1989).

À la suite de DictaMed, à partir de 1993, un second analyseur syntaxique a utilisé la gestion d'hypothèses. Il a été réalisé dans le cadre de la thèse de Pierre-Olivier El Guedj (1996) et il s'est fondé sur des techniques de « charts ». Bien que les terminologies soient différentes, Mats Wirén (1992, pp 121-123) a montré que les charts et les ATMS sont équivalents sur la plupart des points. La seule différence importante est que la gestion des hypothèses d'un chart est « linéarisée » par la structure de gauche à droite des mots d'une phrase.

3.
Analyse syntaxique

3.1 Présentation

J'ai abordé l'analyse syntaxique par le biais du dialogue oral avec une machine. Le premier système de dialogue réalisé avec l'aide d'étudiants – *DictaMed* – utilisait comme notation grammaticale une variante des DCG – *Definite Clause Grammars* (Nugues et al. 1991; Nugues et al. 1992; Nugues et al. 1993(a); Nugues et al. 1993(b); Cazenave et al. 1994).

Le formalisme des DCG (Pereira & Warren 1980) est celui des grammaires syntagmatiques. Il convient pour des dialogues limités, mais il s'avère insuffisant pour une énonciation plus spontanée. Par ailleurs, ce formalisme est incapable de décrire des textes écrits tout-venant. Ceci à cause de l'incorrection de certaines phrases ou des défauts de couverture grammaticale. En admettant qu'on puisse obtenir une couverture exhaustive, la grammaire correspondante impliquerait une complexité d'analyse et une ambiguïté extrêmement importante, impossibles à traiter telles quelles.

Pour apporter une solution, nous avons étudié et réalisé un système d'analyse syntaxique (El Guedj & Nugues 1994, El Guedj & Nugues 1996, El Guedj 1996) se fondant sur la technique des Charts actifs (Earley 1970). Le système fonctionne en mode ascendant et descendant. Il utilise un formalisme dual combinant des grammaires syntagmatiques et de dépendance.

L'analyseur peut gérer les treillis de mots. Ces treillis correspondent à la suite des n meilleures hypothèses de mots que la plupart des systèmes de reconnaissance fournissent pour chaque mot prononcé par un locuteur. Chaque hypothèse correspond à une liste de mots ordonnés et affectés d'un coefficient de mérite – la *N-Best list* –. Notre analyseur examine à chaque fois le premier mot de la liste et peut le rejeter si celui-ci ne mène pas à une expression grammaticale correcte; il considère alors les mots suivants de la liste.

L'analyseur peut procéder à une analyse partielle des constituants et gérer plusieurs hypothèses sur la structure de ces constituants. De cette manière, il peut identifier certaines parties correctes d'une phrase, même si globalement cette phrase est incorrecte ou si elle n'est pas entièrement couverte par la grammaire. L'analyseur retiendra les segments corrects et il laissera les autres de côté. L'analyseur enfin rattache entre eux les segments retenus en utilisant une grammaire de dépendance.

3.2 Pourquoi une analyse syntaxique?

L'amélioration continue des performances des systèmes de reconnaissance vocale ouvre la voie à des applications de dialogue évolué. Le taux de reconnaissance est de plus de 95% de mots correctement reconnus pour des vocabulaires qui atteignent maintenant plusieurs dizaines de milliers de mots. Ces systèmes peuvent opérer à

travers un microphone ou bien à travers un téléphone. Dans ce dernier cas, le vocabulaire n'excède cependant pas cent mots. Ces systèmes permettent, au delà de la simple dictée, la construction de serveurs téléphoniques élaborés et d'interfaces linguistiques interactives.

Parallèlement, l'analyse automatique d'informations et son extraction a connu une explosion. Cette explosion est pour large part due à la disponibilité d'une masse considérable de textes électroniques, notamment à travers l'Internet. Des « moteurs de recherche » sont maintenant capables de collecter des téra-octets de données en quelques semaines. Cette masse de données est d'un potentiel considérable à condition de pouvoir l'exploiter automatiquement. Elle offre la possibilité de mettre à la portée de chacun des connaissances et un savoir à un coût bien moindre qu'aujourd'hui.

Les modèles linguistiques actuels de nombreux systèmes universitaires et industriels, que ce soit pour les systèmes d'extraction d'informations ou bien les machines de reconnaissance vocale, se fondent sur certaines propriétés statistiques des mots ou des suites de mots. On utilise la formule $Tf \times Idf$ – la fréquence des termes d'un document par la fréquence inverse du nombre de documents qui contiennent ce terme – pour la recherche d'informations et les n-grammes pour la reconnaissance vocale. Ces probabilités de séquence peuvent aussi contribuer à déterminer les catégories syntaxiques des mots, mais elles ne permettent pas de saisir le sens de phrases, ni d'un texte complet.

On considère généralement qu'une analyse syntaxique structurelle, de couverture convenable, est nécessaire pour réaliser une interprétation sémantique robuste et pour comprendre un énoncé ou un texte. La structure permet par exemple de déterminer le sujet, le verbe et les compléments d'une phrase. Grâce à une interprétation sémantique, une interface orale pourra exécuter des actions, répondre à des questions, et plus généralement, assurer une interactivité et un dialogue. Un système de recherche d'informations pourra quant à lui, par exemple, distinguer l'affirmation d'un fait de sa négation ou de sa réfutation.

3.3 Représenter la syntaxe

On définit communément la syntaxe comme l'arrangement des mots dans une phrase et dans un discours ainsi que comme les rapports que les mots ont entre eux. Pour ceci, l'un des objectifs de l'analyse syntaxique est d'en décrire la structure. Cependant, il n'y a pas d'accord complet sur ce qu'est précisément une structure syntaxique, ni sur les moyens de la représenter ou de l'analyser. Une tendance assez constante des grammairiens a cependant été le désir de l'isoler des autres parties de la linguistique pour faire du domaine syntaxique un centre autonome. Ce domaine s'insérant entre d'une part les mots et leur morphologie et d'autre part la sémantique. La syntaxe jouerait alors le rôle de nœud d'articulation entre les mots qui sont la matière brute et le sens, qui est « utile ».

Les premières théories grammaticales datent de l'antiquité puis elles se sont perfectionnées au Moyen Âge. Parmi les concepts essentiels associés à l'analyse syntaxique, il y a la classification des mots en un nombre limité de catégories : les

parties du discours et ainsi que les relations et l'influence que les mots d'une phrase ont entre eux : les fonctions. Oswald Ducrot et al. (1995) donne des définitions très claires de ces concepts. Noam Chomsky (1966) retrace l'évolution décisive, au XVII^e et au XVIII^e siècles, vers la grammaire moderne que nous connaissons aujourd'hui. Il rend aussi justice au rôle fondamental joué par les linguistes cartésiens.

La syntaxe a été au cœur de nombreuses théories linguistiques de la deuxième moitié du XX^e siècle. Son succès a été tel qu'elle a un moment éclipsé les écoles linguistiques pour lesquelles elle était moins centrale. Il est assez difficile d'expliquer cette réussite sinon par l'analogie entre la langue et les programmes informatiques qui a inspiré un certain nombre de grammairiens et par la figure exemplaire – et controversée – de Noam Chomsky. Les théories de Chomsky ont dominé le monde de la linguistique de ces quarante dernières années. Ses écrits (1957) – autant que sa personnalité – ont eu une influence inégalée dans les universités d'Europe et d'Amérique du Nord.

Avec l'informatique, le traitement de la langue est devenu une partie du monde de la production industrielle. Ceci a eu des répercussions considérables sur les méthodes de certains linguistes et des ingénieurs de la langue. L'augmentation continue de la puissance des ordinateurs ajoutée à la disponibilité de très grands corpus, d'abord en anglais et maintenant dans beaucoup de langues, permet de valider plus facilement et plus vite une théorie. Ceci évite parfois de poursuivre des voies fausses et au contraire de sélectionner et d'explorer plus rapidement celles qui sont prometteuses. Bien que l'analyse syntaxe de phrases et de textes ne soient par atteinte en totalité, elle fait chaque année des progrès; et, sans doute, des analyseurs décrivant la structure complète de textes verront le jour au début du III^e millénaire.

3.4 Deux paradigmes pour représenter la syntaxe

3.4.1 Les grammaires génératives

En informatique linguistique, le paradigme de représentation le plus largement admis est celui des grammaires génératives. Ces grammaires sont connues pour l'essentiel par les travaux de Chomsky (1957).

Les grammaires génératives – ou formelles – postulent qu'une phrase, appelée l'axiome de la grammaire, peut se décomposer dans une suite d'entités : les constituants immédiats ou les syntagmes. Ces constituants sont eux-mêmes décomposables de la même manière jusqu'à atteindre les mots ou les morphèmes qui constituent la phrase : les éléments terminaux. Il en résulte qu'une langue est engendrée par une grammaire qui consiste en un ensemble de règles de production récursives.

Les grammaires génératives se représentent facilement par un formalisme de règles de réécriture tel que les *Definite Clause Grammars* (Pereira & Warren 1980). On décrit la structure d'un groupe syntaxique non terminal dans une suite de groupes plus simples grâce à des règles – *phrase-structure rules* –. Le Tableau 2 donne un exemple de grammaire pour analyser une phrase élémentaire. Les règles partent de

l'axiome initial (*phr*) et se terminent par des catégories lexicales. Les règles permettent aussi de représenter le vocabulaire terminal. L'analyse syntaxique correspond à la décomposition d'une phrase grâce à ces règles. Elle permet d'en décrire la structure sous la forme d'un ou plusieurs arbres.

<i>phr</i> --> <i>sn</i> , <i>sv</i> .	Une phrase, l'axiome initial, consiste en un syntagme nominal et un syntagme verbal
<i>sn</i> --> <i>art</i> , <i>n</i> .	Un syntagme nominal consiste en un article et un nom
<i>sv</i> --> <i>v</i> , <i>sn</i> .	Un syntagme verbal consiste en un verbe et un syntagme nominal

Tableau 2 Règles syntagmatiques décrivant la structure d'une phrase élémentaire.

3.4.2 Les grammaires de dépendance

Les grammaires de dépendance sont une autre forme de représentation (Tesnière 1959, Mel'cuk 1988). Ces grammaires établissent des liens entre les mots et forment des relations de dépendance. Dans une relation, le mot dominant est appelé le régissant et le mot dominé est le subordonné; un mot pouvant être régissant d'un ou plusieurs mot et à son tour subordonné d'un autre mot.

Une grammaire de dépendance établit un graphe d'« attraction » des mots d'une phrase. Chaque mot a un régissant et un seul, à l'exception de la tête de la phrase, et chaque mot peut avoir zéro, un ou plusieurs subordonnés. Les grammaires de dépendance organisent hiérarchiquement les mots d'une phrase et, dans la plupart des cas, le verbe est la tête de cette hiérarchie. Le verbe domine un ou plusieurs noms, le nom domine son article, ses adjectifs, etc. (Tableau 3).

<i>v</i> <-- <i>n</i> .	Le verbe est le régissant d'un nom
<i>n</i> <-- <i>art</i> .	Le nom est le régissant des autres mots du groupe nominal : ici un article
<i>n</i> <-- <i>adj</i> .	Le nom est le régissant des autres mots du groupe nominal : ici un adjectif

Tableau 3 Règles de dépendance décrivant la structure d'une phrase élémentaire.

À la différence des grammaires syntagmatiques, beaucoup de systèmes utilisant les grammaires de dépendances s'appuient sur le lexique. Les liens de dépendance peuvent être étiquetés et associés à des fonctions. La combinaison du lexique et des fonctions donnent à la sémantique lexicale une position plus importante dans ce type de grammaire. Igor Mel'cuk (1988) affirme que les grammaires de dépendance sont moins sensibles à la rigidité et à la linéarité dans la structure que les grammaires syntagmatiques. De ce fait, elles seraient mieux adaptées pour traiter des langues telles que le latin ou le russe.

Dans leur évolution récente, les grammaires génératives semblent converger vers une intégration des relations de dépendance. Michael Covington (1994) précise les liens entre la théorie X (Chomsky 1970, cité par Stabler 1992) et les grammaires de dépendance. R. Hudson (1984) repris par I. Mel'cuk (1988), affirme que le désir de

distinguer un régissant – un gouverneur – dans une règle de réécriture transforme les règles syntagmatiques en une grammaire de dépendance. Enfin, les formalismes de la fin des années 1980, tels que HPSG (Pollard & Sag 1994), utilisent souvent des règles de réécriture, mais intègrent des relations de dépendance par des traits et des opérations d'unification.

3.5 Des représentations moins profondes

3.5.1 Pourquoi une représentation moins profonde?

Dans de nombreux systèmes actuels de traitement linguistique, on ne met pas directement en œuvre les paradigmes de représentation précédents. En effet, ils sont inutilement lourds pour atteindre certains objectifs pratiques et par ailleurs, ils échouent encore à donner des résultats suffisamment fiables dans le cas général.

Les raisons actuelles de l'échec d'une analyse automatique en profondeur viennent d'abord de l'écriture des règles. Si on arrive assez bien à couvrir un domaine restreint, il est difficile d'écrire à la main une grammaire suffisamment complète pour accepter des gros corpus. Ensuite, en admettant qu'on puisse écrire ces règles, elles résultent inévitablement dans une grammaire importante et complexe à traiter pour les machines actuelles². Enfin, notamment pour les grammaires syntagmatiques, il se pose des problèmes d'ambiguïté difficilement solubles avec de simples considérations syntaxiques. Une phrase pouvant conduire parfois à des dizaines voire des centaines d'interprétations.

Pour obtenir un résultat fiable, on préfère souvent traiter des problèmes jugés plus simples ou moins profonds en comparaison de l'obtention de l'arbre d'analyse complet. On se contente alors d'obtenir des structures partielles ou intermédiaires. Parmi ces analyses moins profondes, l'étiquetage d'une phrase par les parties du discours et la détection des groupes nominaux sont les plus utilisées.

Par ailleurs, dans le cas de l'analyse syntaxique complète, on fait souvent appel à des techniques externes à la grammaire pour résoudre les problèmes d'ambiguïté et de complexité. Ces techniques sont notamment statistiques et heuristiques.

3.5.2 L'étiquetage par les parties du discours

Une des analyses moins profondes les plus souvent pratiquées est l'étiquetage des mots d'une phrase par les parties du discours. Cette analyse – le *Part-Of-Speech Tagging* de la littérature américaine – a été entièrement renouvelée par K. Church en 1988. L'étiquetage est éventuellement précédé d'une lemmatisation. Nous ne rentrerons pas dans le détail de l'analyse morphologique flexionnelle. On peut considérer que ce point est en grande partie résolu par les systèmes de transduction à deux niveaux tels

² Des calculs ont été réalisés pour l'anglais sur le corpus parenthésé Penn TreeBank II, comportant 1.200.000 mots. Il en résulte une grammaire de 17.540 règles pour couvrir la totalité du corpus! Cependant 240 règles suffisent à couvrir 80% du corpus (Gaizauskas et al. 1996).

que KIMMO (Koskenniemi 1983). Pour une présentation et une discussion critique de la morphologie à deux niveaux, on pourra consulter l'article de B. Fradin (1994).

L'étiquetage vise à éliminer l'ambiguïté de classification des mots. Il repose sur l'application de méthodes statistiques ou de règles symboliques pour affecter une catégorie lexicale unique à chacun des mots (Tableau 4). Des méthodes statistiques et symboliques ont été proposées.

The/ART chair/NOUN will/MODAL table/NOUN the/ART motion/NOUN

Tableau 4 L'annotation des mots par des parties du discours (Church & Mercer 1993).

Les méthodes statistiques modélisent l'étiquetage comme des parties du discours traversant un canal bruité et se trouvant transformées en mots (Church 1988, Church & Mercer 1993, Merialdo 1994). On retrouve la chaîne optimale de parties du discours en appliquant la formule de Shannon (Équation 2). Dans cette équation, P désigne la suite de parties du discours et W la suite de mots.

$$\hat{P} = \operatorname{argm} \operatorname{ax} \Pr(P) \cdot \Pr(W | P)$$

Équation 2 La formule de Shannon appliquée aux parties du discours.

Cette formule est approximée en considérant, non pas la chaîne complète, mais des trigrammes de parties du discours ainsi que des couples (mot, partie du discours) (Équation 3, Équation 4) (Merialdo 1994). Les estimations des probabilités s'obtiennent par des statistiques lexicales provenant de grands corpus annotés. La séquence optimale de partie du discours correspond à la valeur maximale de l'Équation 2.

$$\Pr(P) = \Pr(P_1, P_2, \dots, P_n) \approx \prod_{i=1}^n \Pr(P_i | P_{i-1}, P_{i-2})$$

Équation 3 Les simplifications courantes : la probabilité sur la suite de parties du discours se simplifie en un produit de probabilités sur les trigrammes.

$$\Pr(W | P) = \Pr(W_1, W_2, \dots, W_n | P_1, P_2, \dots, P_n) \approx \prod_{i=1}^n \Pr(W_i | P_i)$$

Équation 4 Les simplifications courantes : la probabilité de la suite de mots connaissant la suite de parties du discours se simplifie en un produit de probabilités sur les couples mot, partie du discours.

Les méthodes symboliques utilisent des règles pour étiqueter les mots. Ces méthodes sont dues essentiellement à E. Brill (1995). L'algorithme réalise un premier étiquetage de chacun des mots de la phrase par sa partie du discours la plus fréquente. Il en résulte bien sûr un nombre d'erreurs assez important. L'algorithme change ensuite la partie du discours de certains mots en appliquant séquentiellement un ensemble de règles. Les règles suivent des modèles très simples et tiennent

compte du contexte immédiat à gauche et à droite du mot qu'elles considèrent. Un exemple de modèle de règle appliquée sur un mot i est :

CHANGER ÉTIQUETTE DE MOT _{i} DE E1 EN E2
 SI
 PARTIE DU DISCOURS DE MOT _{$i-1$} EST E3

Les règles d'étiquetage sont apprises automatiquement à partir d'un corpus annoté à la main. L'algorithme d'apprentissage réalise une annotation initiale des mots du corpus en utilisant pour chaque mot considéré sa partie du discours la plus fréquente. Il instancie ensuite toutes les règles pour chaque erreur relevée par rapport à l'annotation manuelle de référence. L'algorithme d'apprentissage retient la règle qui apporte la plus grande réduction d'erreurs. Il modifie l'annotation précédente en appliquant cette règle puis il recommence les étapes précédentes jusqu'à ce que le taux d'erreur soit suffisamment bas. Le Tableau 5 donne les trois premières règles pour l'anglais.

Ancienne étiquette	Contexte	Nouvelle étiquette
NOUN	l'étiquette précédente est TO	VERB, BASE
VERB, PRESENT	une des trois étiquettes précédentes est MODAL	VERB, BASE
NOUN	une des trois étiquettes précédentes est MODAL	VERB, BASE

Tableau 5 Des exemples de règles tirées de Brill (1995).

Comme pour les méthodes statistiques, l'apprentissage automatique des règles repose sur des corpus annotés. C'est bien souvent le Penn Treebank pour l'anglais (Marcus et al. 1993). Ces deux techniques d'étiquetage sont relativement bien maîtrisées et elles atteignent, selon leurs auteurs, des taux d'étiquetage correct supérieurs à 95% pour l'anglais.

3.5.3 La détection des groupes

La détection de groupes réalise le découpage des phrases en groupes nominaux non récursifs et en groupes verbaux. Elle est le prolongement de l'analyse précédente et elle peut utiliser les mêmes techniques. Le découpage en groupes se voit alors comme une affectation de parenthèses ouvrantes et fermantes entre les groupes de mots. Il s'agit donc d'un étiquetage où les étiquettes sont affectées cette fois-ci aux intervalles entre les mots (Tableau 6).

[_NThe government_N] has other [_Nagencies and instruments_N] for pursuing [_Nthese objectives_N]

Tableau 6 Le découpage des groupes nominaux (Ramshaw & Marcus 1995).

Si on ne détecte que les groupes nominaux, le jeu d'étiquettes pour marquer les intervalles peut se réduire à : Intérieur (*in*), Extérieur (*out*) et Entre (*between*). Si on affecte les étiquettes aux mots :

- l'étiquette Intérieur signale que le mot fait partie d'un groupe nominal;
- l'étiquette Extérieur signale que le mot ne fait pas partie d'un groupe;
- l'étiquette Entre signale que le mot est à la fin d'un groupe et que ce groupe est suivi d'un autre immédiatement après.

La détection de groupes fait appel à des techniques statistiques (Church 1988), symboliques (Ramshaw & Marcus 1995), ou bien à des cascades d'automates (Abney 1996). Dans le cas de l'analyse statistique, on maximise les probabilités d'étiquetage des intervalles – gaps – connaissant les mots qui précède et qui suivent l'intervalle ainsi que les parties du discours de ces mots (Équation 5). On peut simplifier l'équation en ne considérant que les parties du discours.

$$\prod_{i=2..n} \hat{P}(G_i | w_{i-1}, t_{i-1}, w_i, t_i)$$

Équation 5 La formule d'étiquetage des intervalles où w_i est le mot d'indice i et t_i sa partie du discours. G_i est la valeur d'étiquetage de l'intervalle.

Dans le cas d'une détection symbolique, on réalise un premier étiquetage des intervalles de mots. On balaye ensuite les étiquettes et on les modifie éventuellement par des règles. Ces règles prennent en compte le contexte immédiat de l'étiquette. Il est constitué :

- des étiquettes de découpage de groupes (sur une étendue de deux mots à gauche et à droite);
- des mots et des parties du discours (sur une étendue de trois mots à gauche et à droite).

Des exemples de telles règles sont donnés sur le Tableau 7.

Ancienne étiquette	Contexte	Nouvelle étiquette
I	$T_1=O, P_0=JJ$	O
-	$T_{-2}=I, T_{-1}=I, P_0=DT$	B
I	$T_{-1}=I, W_0=whO$	B

Tableau 7 Des exemples de règles tirées de Ramshaw & Marcus (1995).

I = *in*, O = *out* et B = *between*. -1, 0 et 1 sont les indices respectivement du mot précédent, du mot courant et du mot suivant. P désigne la partie du discours, T l'étiquette de l'intervalle, W le mot. JJ désigne un adjectif et DET un déterminant.

Les auteurs des techniques d'étiquetage de groupes nominaux rapportent des taux de détection correcte de l'ordre de 90%

Les représentations syntaxiques « plus faibles » permettent de résoudre un certain nombre de problèmes très concrets de l'informatique linguistique. Par exemple pour la synthèse vocale, l'étiquetage d'un mot par sa catégorie élimine un grand nombre d'ambiguïtés de transcription de graphèmes en phonèmes. La détection des groupes permet d'aborder le domaine des liaisons du français et de la prosodie. Il trouve aussi des applications dans le domaine de la recherche documentaire et de l'indexation. Enfin, les deux techniques d'étiquetage, des mots et des groupes, peuvent s'utiliser comme un premier niveau pour construire l'arbre syntaxique complet de la phrase.

3.6 Un système stochastique d'analyse combinant plusieurs paradigmes

3.6.1 Introduction

Dans un article récent, Michael Collins (1996) a décrit un mécanisme d'analyse syntaxique reposant sur une structure en couches. Il utilise l'étiquetage, la détection de groupes ainsi que des statistiques sur les bigrammes de mots rattachés par une relation de dépendance. L'algorithme repose sur une démarche ascendante. Après avoir segmenté les groupes de mots, il fait appel une grammaire de dépendance pour relier les groupes entre eux et il construit l'arbre syntaxique à partir de ces dépendances.

Nous présentons ce système car il permet d'obtenir les meilleurs résultats publiés à ce jour pour l'anglais. Il retrouve plus de 85% de structures syntaxiques sur le corpus du *Wall Street Journal*. Il dépasse ainsi l'analyseur statistique SPATTER (Magerman 1994) qui était le précédent tenant du titre.

3.6.2 Correspondance entre le modèle syntagmatique et les grammaires de dépendance

La méthode de M. Collins repose sur le lien entre le modèle syntagmatique et les grammaires de dépendance. Elle fait systématiquement correspondre un graphe de dépendances à une structure d'arbre. Pour ceci, la méthode se fonde sur le fait que chaque syntagme contient un régissant. Elle identifie alors dans chaque règle de réécriture un régissant dans sa partie droite. Par exemple dans la règle

$$sn \rightarrow art, \underline{n}, adj.$$

n est régissant.

Pour établir la correspondance entre les deux modèles, on propage les régissants de manière ascendante : la partie gauche ayant comme régissant, le régissant de la partie droite. On procède de la même manière pour les toutes règles du modèle

syntagmatique et on fixe un régissant parmi les constituants de la partie droite. Ainsi, dans la règle :

$$\text{phr} \rightarrow \text{sn}, \underline{\text{sv}}.$$

sv est régissant. De cette façon, en utilisant les deux règles syntagmatiques précédentes, on établira la dépendance entre le nom du sujet et le verbe du prédicat.

Avec cette technique, l'arbre syntaxique d'une phrase peut se transformer en un graphe de dépendance. Le graphe se représente par la séquence

$$D = \{\text{Rég}(1), \text{Rég}(2), \dots, \text{Rég}(n)\}$$

où chaque mot de position i est relié à son régissant : $\text{Rég}(i)$. Le régissant est défini par son indice de position dans la phrase et par un triplet caractérisant la règle syntagmatique qui relie les deux mots. Cette règle est celle permettant de passer de la dépendance au syntagme. Le triplet est formé de :

1. la catégorie dont provient le subordonné dans la partie droite de la règle;
2. la catégorie de la partie gauche;
3. la catégorie dont provient le régissant dans la partie droite de la règle.

Dans le cas de la dépendance sujet-verbe, la règle syntagmatique en jeu est $\text{phr} \rightarrow \text{sn}, \text{sv}$. Le triplet $\langle \text{sn}, \text{phr}, \text{sv} \rangle$ représente donc cette relation. La transformation d'un arbre syntaxique complet résulte dans une suite de relations de dépendance (Figure 10).

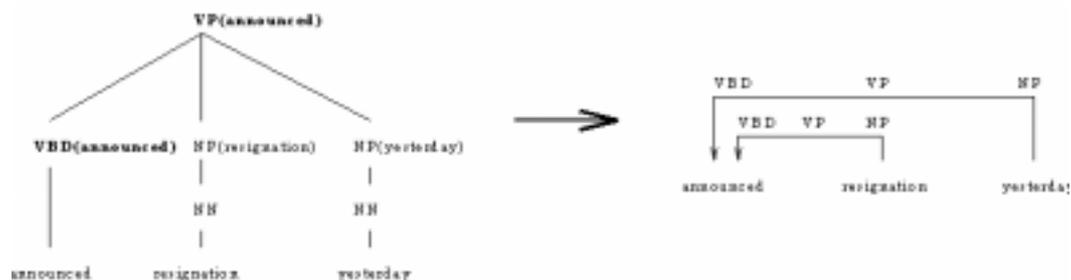


Figure 10 La transformation d'un arbre syntaxique en relations de dépendance (Collins 1996).

3.6.3 Le calcul des relations de dépendance

L'algorithme de Michael Collins utilise la probabilité d'un rattachement de deux mots par une relation de dépendance. L'estimation des probabilités se fait par des statistiques sur le corpus complètement parenthésé du Penn TreeBank (Marcus et al. 1993).

Les statistiques de dépendance sont obtenues en transformant chaque arbre syntaxique du corpus. Ceci permet de mesurer le nombre de relations de dépendance R reliant deux mots w_i et w_j avec les parties du discours t_i et t_j , sur le nombre total d'apparitions de ces deux mots, avec les mêmes parties du discours, dans une même phrase (Équation 6). Dans le calcul des relations, les groupes nominaux sont réduits au seul nom.

$$F(R | \langle w_i, t_i \rangle, \langle w_j, t_j \rangle) = \frac{C(R, \langle w_i, t_i \rangle, \langle w_j, t_j \rangle)}{C(\langle w_i, t_i \rangle, \langle w_j, t_j \rangle)}$$

Équation 6 L'estimation d'une relation de dépendance R entre les mots w_i et w_j dont les parties du discours sont t_i et t_j .

Les dépendances sont évaluées de façon stochastique. La détermination des liens entre les mots d'une phrase se fait en maximisant la formule de l'Équation 7.

$$\prod_{j=1}^m F(R_j | \langle w_j, t_j \rangle, \langle w_{h_j}, t_{h_j} \rangle)$$

Équation 7 Le calcul des dépendances. Où m est le nombre de mots de la phrase. F représente la probabilité qu'il existe une dépendance R entre le mot w_j d'indice j et de partie du discours t_j et celui w_{h_j} d'indice h_j .

3.6.4 L'analyseur syntaxique

L'analyse syntaxique consiste à construire les arbres syntaxiques en partant des relations de dépendance. Elle cherche à maximiser au cours de l'analyse la probabilité de découpage de la phrase en groupes nominaux par la probabilité de liens de dépendance entre les groupes.

L'analyse syntaxique repose sur une méthode ascendante en deux étapes. La première étape réalise l'étiquetage des mots par des parties du discours. Cet étiqueteur est indépendant du reste du système et fonctionne, d'après ses auteurs, avec une précision de 97%. La deuxième étape intègre la détection de groupes et l'établissement des dépendances dans une analyse par un chart ascendant.

Le chart contient les découpages possibles d'une phrase en groupes nominaux non récursifs. Les découpages se fondent sur une méthode stochastique. Elle calcule la probabilité d'étiquetage des intervalles – des gaps – en tenant compte des mots, des parties du discours qui suivent et précèdent, ainsi que de la présence d'une virgule. La probabilité d'une séquence de groupes nominaux est donnée par l'Équation 8. C'est la même que celle de l'Équation 5 avec la prise en compte de la présence d'une virgule c_i .

$$\prod_{i=2..n} \hat{P}(G_i | w_{i-1}, t_{i-1}, w_i, t_i, c_i)$$

Équation 8 La probabilité d'une séquence de groupes nominaux.

Les groupes nominaux sont réduits au seul nom qui les compose. On ne tient pas compte des articles, des adverbes et des adjectifs qui précèdent le nom principal. Ils se rattachent sans ambiguïté au nom. Le chart stocke donc les groupes nominaux possibles et les autres mots munis chacun d'une probabilité de découpage.

L'analyse combine deux constituants adjacents, C_1 et C_2 en un constituant supérieur C_3 en le munissant d'une nouvelle probabilité. Pour être combinés, les deux constituants doivent être séparés par l'étiquette de séparation des groupes de mots : $\text{Gap} = S$. Pour tenir compte de la correspondance des deux modèles syntaxiques, ils doivent aussi être reliés par une relation de dépendance. Par ailleurs, l'algorithme empêche les croisements d'arc.

La probabilité du nouveau constituant est déterminée par le produit des probabilités de C_1 , C_2 , par la probabilité d'avoir une séparation entre les deux constituants et par la probabilité de dépendance entre les deux (Équation 9).

$$S(C_3) = S(C_1) \times S(C_2) \times \Pr(\text{Gap} = S | C_1, C_2) \times \Pr(\langle C_1, C_3, C_2 \rangle | C_1, C_2)$$

Équation 9 La probabilité d'un nouveau constituant.

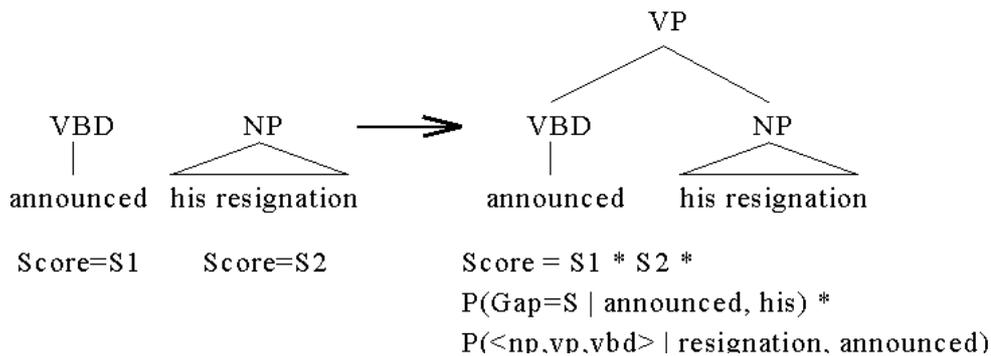


Figure 11 Analyse syntaxique (Collins 1996).

Dans cette formule, on ne tient pas compte des positions des mots, ni de leur ordre, ni de leur distance. Michael Collins ajoute un terme représentant une distance heuristique qui n'est pas détaillée dans son texte.

3.6.5 Discussion

Le système de Michael Collins est à ce jour celui qui obtient les meilleurs résultats publiés pour l'anglais. Pour l'évaluer, il compare les constituants retrouvés par l'analyseur à ceux de l'annotation manuelle. L'évaluation utilise les critères de rappel (recall) et de précision classiques dans le domaine de la recherche d'informations. Le rappel se définit comme le nombre de constituants corrects retrouvés par l'analyseur sur le nombre de constituants de l'annotation manuelle. La précision se définit comme le nombre de constituants corrects retrouvés par l'analyse sur le nombre total

de constituants retrouvés. L'analyseur obtient des taux de précision et de rappel d'environ 85%

L'analyseur de Collins est original car il utilise un paradigme de traitement mêlant les grammaires syntagmatiques et les grammaires de dépendance. Il produit un arbre syntagmatique comme représentation finale. L'analyseur procède en deux étapes principales qui sont la détection des groupes nominaux et verbaux et l'établissement de relations de dépendance entre les groupes. Il utilise un chart pour stocker les hypothèses et sélectionne les groupes et les liens en fonction du critère de maximum de vraisemblance. Les statistiques sont tirées de corpus annotés à la main.

Nous avons présenté le système d'analyse syntaxique de Michael Collins car il est voisin, dans ses concepts et sa structure, de celui que nous avons construit pour le français (El Guedj & Nugues 1994, El Guedj & Nugues 1996, El Guedj 1996). À la différence de Michael Collins, notre système procède avec des règles symboliques au lieu de critères statistiques. Les méthodes symbolique et statistique sont parallèles et leur application à l'annotation par les parties du discours montre qu'elles ont un potentiel semblable quant aux résultats.

3.7 Un système symbolique d'analyse combinant plusieurs paradigmes

3.7.1 L'analyse syntaxique et l'anglais

La plupart des systèmes d'analyse décrits dans la littérature et qui atteignent des taux de succès élevés opèrent sur l'anglais. Or cette langue possède certaines particularités qui rendent les formalismes valides pour elle, parfois difficile à transposer pour les autres (Mel'cuk 1988) :

- l'anglais a une morphologie flexionnelle très pauvre;
- l'anglais a un ordre des mots très rigide.

Dans son argumentation, Igor Mel'cuk (1988) se fait l'avocat des grammaires de dépendance. Il affirme notamment que les grammaires syntagmatiques n'auraient pas pu être inventées par un Russe ou par un locuteur latin. En revanche, les dépendances permettent de rendre compte de la mobilité des mots et de la discontinuité des constituants. (Igor Mel'cuk emploie le mot « groupements »). Nous pensons que ceci est aussi valide, dans une moindre mesure, pour le français.

Nous avons adapté les deux paradigmes en les intégrant dans un même système symbolique. Ce système utilise une grammaire syntagmatique (El Guedj & Nugues 1994) ainsi qu'une grammaire de dépendance (El Guedj & Nugues 1996, El Guedj 1996) et il combine les deux représentations syntaxiques.

Le système d'analyse permet de retenir la compacité de description des grammaires syntagmatiques à un premier niveau. Ce premier niveau correspond aux groupes nominaux et aux groupes verbaux qui ont une structure relativement fixe en

français. L'analyseur utilise ensuite des règles de dépendance pour rattacher les groupes entre eux. Il permet ainsi de s'affranchir de leur mobilité dans la phrase.

La conception du système a été incrémentale autour d'un noyau d'unification. Son développement s'est fait en deux temps. Le premier analyseur a été conçu en respectant la structure classique d'un chart (Earley 1970, Younger 1967) et en la munissant d'extensions. Ce premier chart utilise des grammaires syntagmatiques munies de traits.

Le second système reprend les éléments du premier analyseur en se fondant sur la structure du chart précédent pour détecter des segments nominaux et verbaux. Il complète cette structure pour procéder à l'analyse de plus haut niveau. Il utilise une grammaire syntagmatique, des règles de précédence et une grammaire de dépendances :

1. la grammaire syntagmatique réalise un découpage de la phrase en segments pour donner un treillis de groupes syntaxiques nominaux, verbaux, ... Les segments produits sont en surnombre par rapport aux groupes réels de la phrase;
2. les règles de précédence considèrent les bigrammes formés par les groupes syntaxiques. Elles vérifient qu'un segment à un prédécesseur licite et l'élimine sinon. Ces règles permettent d'enlever une partie des éléments du treillis qui sont incohérents avec le reste de la phrase;
3. les règles de dépendances relient les segments entre eux et construisent une représentation structurelle de la phrase.

3.7.2 Les étapes de l'expérimentation

Le premier système et sa grammaire ont été conçus en considérant des corpus de dialogues oraux. Ces dialogues mettaient en jeu un utilisateur et un agent conversationnel placés dans un monde virtuel. Le cadre du dialogue était délimité par un certain nombre d'interactions possibles entre l'utilisateur et le monde virtuel (Godéreaux 1997).

Ce premier analyseur a permis d'implanter un agent qui déplace l'utilisateur en son nom dans un monde virtuel. L'analyseur a été muni de mécanismes destinés à traiter l'oral. Il produit avec succès des arbres syntaxiques pour des énoncés relativement courts, souvent à l'impératif. Les énoncés correspondent au dialogue qu'un utilisateur peut avoir pour commander sa navigation dans un environnement virtuel. Nous décrivons plus complètement l'agent conversationnel *Ulysse* dans le chapitre suivant.

Nous avons ensuite considéré un corpus de textes médicaux. Ce corpus correspond à des comptes rendus de consultation. L'ensemble des comptes rendus a été produit par plusieurs praticiens. Il contient des phrases longues – ou courtes – d'un style extrêmement varié. Malgré un certain nombre d'essais, il nous a été impossible d'écrire des grammaires syntagmatiques pour ces énoncés. Nous avons conclu à l'inadéquation des grammaires formelles pour analyser ce type de phrases. Ceci nous a amenés à réviser l'architecture du système pour le faire fonctionner différemment.

Le modèle d'analyse du second prototype se fonde sur l'observation suivante en français. Il existe des segments dont l'ordre est relativement rigide. Ces segments sont essentiellement les groupes nominaux non récursifs et les groupes verbaux.

Dans ces groupes, la liberté est restreinte. On peut éventuellement inverser un adjectif et un nom dans un groupe nominal, mais pas un article et un nom.

En revanche, à la différence de l'anglais, certains groupes sont plus mobiles. La structure de ces groupes dans la phrase s'analyse difficilement avec des grammaires syntagmatiques qui lui imposent des contraintes de linéarité. On pourrait bien écrire de manière combinatoire des règles, mais elles multiplieraient l'ambiguïté. Les relations des groupes avec les autres mots de la phrase sont plus apparentes si on les explique par une dépendance fonctionnelle. Nous avons choisi de les analyser avec le formalisme des grammaires de dépendance.

3.7.3 La structure d'un chart

Un chart repose sur une représentation où les intervalles – les nœuds – entre les mots sont numérotés de 0 à n . Les nœuds sont reliés par des arcs qui définissent un constituant. Le constituant peut être complètement analysé ou bien en cours d'analyse.

De manière classique, on représente l'état d'avancée de l'analyse d'un constituant dans un chart par un point (\bullet). Ainsi

gn art nom \bullet

est un groupe nominal complètement analysé et constitué d'un article et d'un nom. Quant à

gn art \bullet nom

elle représente un groupe nominal en cours d'analyse. On a trouvé un article et on cherche un nom pour le compléter. La Figure 12 représente le chart du groupe nominal *la table jaune* avec les règles qui permettent de l'analyser.

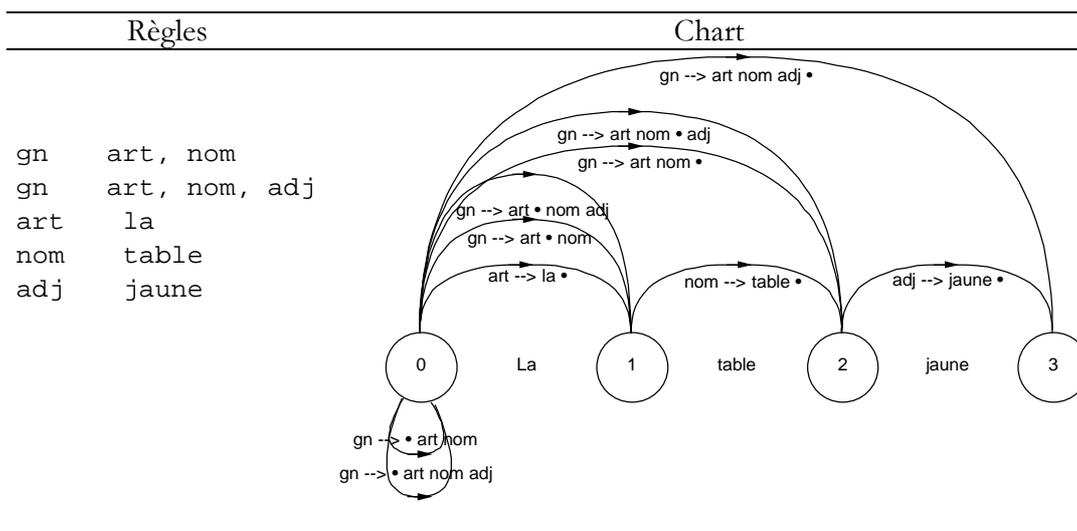


Figure 12 Des règles de grammaire et le chart correspondant.

Un chart permet les analyses ascendantes et descendantes. Dans le mode ascendant aussi bien que descendant, le chart mémorise les arcs en cours d'analyse – les arcs actifs –. Une fois qu'un constituant vient d'être complètement analysé – l'arc est alors inactif –, on le stocke dans un agenda jusqu'à ce qu'il soit lui-même inséré dans le chart.

Le mode ascendant part des mots d'entrée et déclenche les règles pour construire la structure complète en remontant. Il applique la règle ascendante pour étiqueter le mot et l'introduire dans le chart (Tableau 8) et la règle fondamentale pour faire avancer le point d'analyse des arcs actifs (Tableau 9).

Tant qu'il y a des mots :

1. Si l'agenda est vide, accepter un mot et l'ajouter dans l'agenda avec sa catégorie autant de fois qu'il a d'interprétations : $C \quad \text{Mot} \bullet$
 2. Extraire un constituant C de l'agenda (on note sa position de i à j).
 3. Pour chaque règle de la forme $B \quad C \ C_1 \ C_2 \dots \ C_n$, ajouter un arc actif $B \bullet C \ C_1 \ C_2 \dots \ C_n$, dans le chart de i à j
 4. Ajouter C dans le chart en utilisant la règle fondamentale
-

Tableau 8 La règle ascendante.

Ajout d'un constituant C entre les positions j et k :

1. Ajouter C entre j et k dans le chart.
 2. Pour tous les arcs actifs $B \quad C_1 \ C_2 \dots \ C \dots \ C_n$, de i à j , ajouter le nouvel arc actif $B \quad C_1 \ C_2 \dots \ C \bullet \dots \ C_n$ de j à k .
 3. Pour tous les arcs actifs $B \quad C_1 \ C_2 \dots \ C_n \bullet C$, de i à j , ajouter le constituant B de i à k dans l'agenda.
-

Tableau 9 La règle fondamentale.

Le mode descendant démarre de la structure à analyser – la phrase – pour descendre jusqu'aux mots. Ce mode est aussi connu sous le nom d'algorithme de Earley (1970). L'algorithme est constitué de trois modules qui s'enchaînent :

- Le premier module, le prédicteur, recherche les règles qui peuvent faire avancer les arcs actifs. Pour ceci, pour tous les arcs $B \quad C_1 \ C_2 \dots \ C \dots \ C_n$, on recherche récursivement les règles de la forme $C \quad X_1 \dots \ X_k$ que l'on introduit dans le chart sous la forme $C \bullet X_1 \dots \ X_k$
- Le second module, l'accepteur (ou le scanner), accepte un nouveau mot de l'entrée. Il l'introduit dans l'agenda avec ses interprétations possibles : $C \quad \text{Mot} \bullet$.
- Le troisième module, le compléteur, ajoute chaque constituant C de l'agenda dans le chart. Il utilise la règle fondamentale pour le combiner avec les arcs actifs. Il détermine quels sont les constituants de plus haut niveau terminés. Il les ajoute dans l'agenda.

On initialise cet algorithme avec l'entrée $\bullet S$ où S est l'axiome de la grammaire. C'est souvent la phrase ou, pour la grammaire précédente, gn .

3.7.4 Le premier système et ses extensions

L'architecture du premier analyseur se fonde sur la structure d'un chart classique. Le système utilise des règles syntagmatiques. Il comprend un noyau d'unification qui permet d'enrichir les règles par des traits (Sabah 1990). Les équations d'unification utilisent le format PATR (Shieber 1986). Le chart fonctionne en mode ascendant et en mode descendant.

L'algorithme d'analyse a été doté d'extensions pour gérer une entrée vocale en mode ascendant. Ces extensions ont pour but d'accepter et de traiter les mots un par un. Ceci à la différence d'un chart classique qui traite des phrases complètes et stables. Le traitement incrémental permet d'exploiter les hypothèses de reconnaissance d'une carte vocale. En effet, à l'énoncé d'un mot, la plupart des systèmes de reconnaissance vocale peuvent retourner, non pas un mot, mais la liste de N meilleures hypothèses du mot qu'ils croient avoir reconnu (la N -Best List). Le traitement incrémental permet aussi dans une certaine mesure d'émettre des hypothèses sur un mot manquant.

Pour implanter les extensions en mode ascendant, nous avons défini une nouvelle règle : la règle d'addition de mot (Tableau 10). Cette règle se substitue à la règle ascendante (El Guedj 1996). À la réception d'un mot, la règle d'addition de mot génère récursivement tous les arcs actifs possibles partant de ce mot. Elle vérifie que le mot pourrait *in fine* s'intégrer dans une phrase complète.

Fondamentalement, la règle d'addition de mot permet de vérifier si le dernier mot accepté peut former avec les mots déjà énoncés, un début de phrase correcte. En complétant les règles de la Figure 12 par celle du Tableau 11, on obtient le chart de la Figure 13 lorsqu'on applique la règle ascendante. La règle d'addition de mots produit le chart de la Figure 14.

Lors de l'acceptation d'un nouveau mot :

1. Ajouter le mot dans l'agenda avec sa catégorie autant de fois qu'il a d'interprétations : $C \quad \text{Mot} \bullet$
 2. Extraire un constituant C dans l'agenda (on note sa position de i à j).
 3. Pour chaque règle de la forme $B \quad C \ C_1 \ C_2 \dots \ C_n$, ajouter un arc actif $B \bullet C \ C_1 \ C_2 \dots \ C_n$, dans le chart de i à j
 4. Appliquer récursivement le point 3 à toutes les règles de la forme $A \quad B \ B_1 \dots \ B_n$
-

Tableau 10 La règle d'addition de mot.

```

phr --> sujet pred
pred --> verbe objet
objet --> gn
sujet --> gn

```

Tableau 11 Complément de règles.

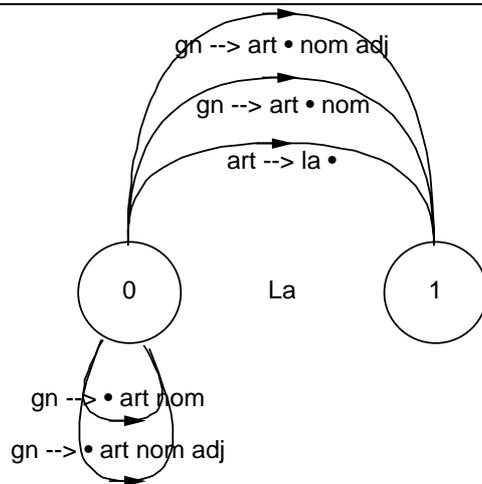


Figure 13 État du chart avec la règle ascendante classique.

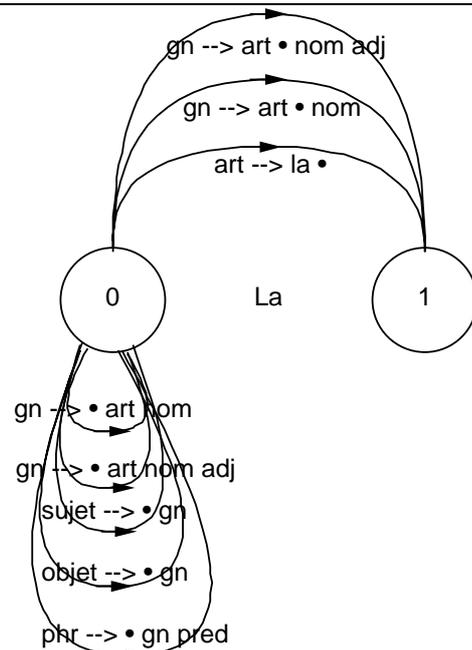


Figure 14 État du chart avec la règle d'addition de mot.

Avec la règle d'addition de mot, l'analyseur peut réaliser un contrôle de la syntaxe en même temps qu'il reçoit les mots. Notamment, si le mot courant est incorrect, c'est-à-dire qu'aucune règle ne l'accepte pour former une phrase, on peut le retirer et prendre le suivant dans la liste fournie par le système de reconnaissance. On peut aussi décider de le rejeter purement et simplement et passer au suivant.

Le chart peut aussi gérer un ou plusieurs mots manquants. Pour ceci, on utilise des mots jokers dotés de toutes les parties du discours. On peut les introduire dans la phrase de façon à ce qu'elle constitue une chaîne correcte. On les place entre tous les mots acceptés. Pour vérifier qu'une phrase est correcte, on tente de constituer une suite transitive d'arcs et si on ne peut pas, on fait rentrer en jeu les jokers.

Enfin, le chart gère les mots inconnus en leur donnant la catégorie nom. De tels mots peuvent être émis par le système de reconnaissance, car les lexiques de l'analyseur syntaxique et de la reconnaissance vocale sont différents.

3.7.5 L'architecture du second système

Le second système a réutilisé en partie l'architecture du premier et il a intégré une grammaire de dépendance. Il se structure en couches hiérarchiques. Chacune des couches correspond à un niveau de traitement :

- la première couche se charge de la détection des groupes syntaxiques non récursifs – les segments –. Cette couche génère toutes les hypothèses de segments que lui permet les règles. Naturellement, cette partie détecte plus de segments qu'il n'y en a réellement;

- la deuxième couche réduit l'ambiguïté en éliminant une partie des segments suivant des critères de compatibilité avec le contexte immédiat du groupe;
- la troisième couche établit des liens de dépendance entre les segments.

Le système s'intègre dans un chart ascendant. Chacune des couches opère sur les arcs en appliquant des règles pour produire un nouvel ensemble d'arcs. La première couche applique des règles de réécriture syntagmatiques. La seconde vérifie des compatibilités de bigrammes. La troisième utilise des règles de dépendance. Ainsi, il en résulte d'abord des arcs lexicaux, puis des groupes syntaxiques, des groupes syntaxiques valides et enfin des arcs de dépendance (Figure 15).

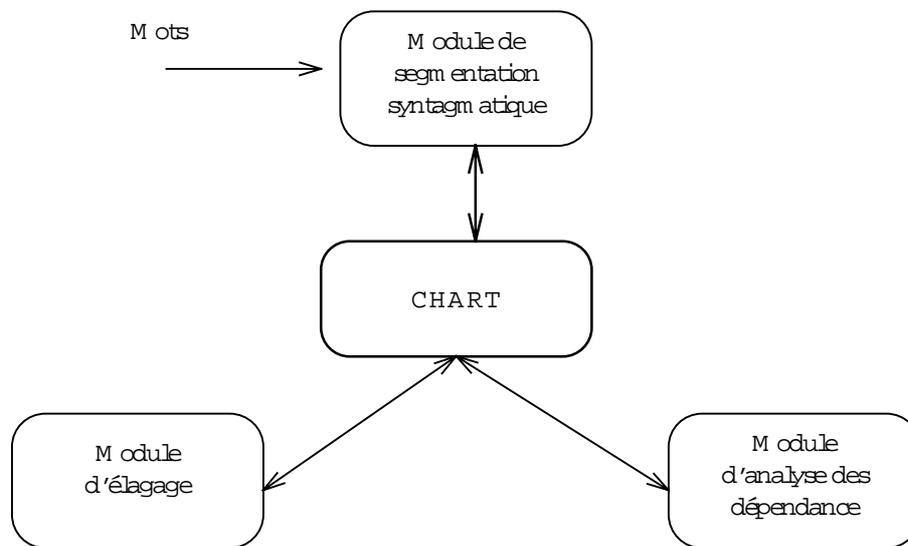


Figure 15 L'architecture du système d'analyse syntaxique.

3.7.6 Les catégories des segments

La structure des segments se fonde sur des règles syntagmatiques. Les règles définissent des groupes syntaxiques non récursifs. Les catégories de ces groupes correspondent à celles de la grammaire classique. Elles sont plus nombreuses que celles utilisées par Church (1988) et Ramshaw & Marcus (1995). Ces auteurs ne considèrent que les groupes nominaux et éventuellement verbaux.

Les différents groupes que nous avons utilisés représentent le niveau syntagmatique de base et servent aussi à déterminer la compatibilité des bigrammes et dans l'établissement des dépendances. Il sont, de façon résumée :

- les groupes nominaux. Ils sont définis comme les groupes de mots formés autour d'un nom. Ils sont composés notamment de déterminants, des adjectifs antéposés, d'un nom et de la plupart des adjectifs postposés. Un adverbe peut précéder les adjectifs;
- les groupes verbaux. Ils sont définis comme les groupes de mots formés autour d'un verbe. Ils sont constitués d'adverbes de négation, de pronoms clitiques,

d'adverbes, d'auxiliaires, d'un verbe à un temps simple ou composé, du pronom personnel sujet s'il y a inversion du sujet;

- les groupes prépositionnels, les adverbes et les locutions adverbiales. Cet ensemble est réuni sous le nom de « compléments ». Les groupes prépositionnels sont définis comme des groupes nominaux précédés d'une préposition;
- les groupes d'adjectifs. Ils sont constitués des segments d'adjectifs attributs et de certains adjectifs postposés;
- les participes passés;
- les prépositions suivies d'infinitifs, comme *pour prendre*, *de faire*.
- les pronoms personnels;
- des entités de structures : les conjonctions, les pronoms relatifs et la ponctuation.

Chacun des groupes est défini par un ensemble de règles de structure syntagmatique avec des traits. Les traits correspondent notamment à des caractéristiques d'accord et de sémantique. Les traits sont appliqués comme des contraintes d'unification. Le formalisme retenu est celui de PATR.

Le lexique classe les mots suivant des parties du discours semblables à celles de MULTTEXT (Ide & Véronis 1995, Véronis & Khoury 1995). Les mots comprennent un trait d'accord et un trait sémantique. Par exemple, pour le nom, le trait d'accord comprend les sous traits de genre et de nombre. Le trait sémantique peut notamment prendre les valeurs : temps, lieu, objet, agent, etc. Les traits sont passés au groupe tout entier par unification. Ils permettent notamment de réaliser un étiquetage sémantique des groupes nominaux.

3.7.7 Le découpage des segments

Le découpage de segments s'opère de manière symbolique à l'intérieur d'un chart. Il procède de gauche à droite en acceptant et en analysant les mots un par un. Comme dans un chart classique, les espaces entre les mots forment les nœuds. Ces nœuds sont reliés par des arcs. Les arcs ont comme origine, le début du groupe en cours d'analyse et comme fin, le dernier mot analysé. Ils sont annotés par la règle syntagmatique en cours d'application.

À la différence de l'algorithme classique du chart, l'analyse des groupes opère uniquement sur une fenêtre. Le nœud de départ est appelé le point d'ancrage. Pour être un point d'ancrage, un nœud doit être relié transitivement par une suite d'arcs inactifs partant du nœud 0 (Figure 16).

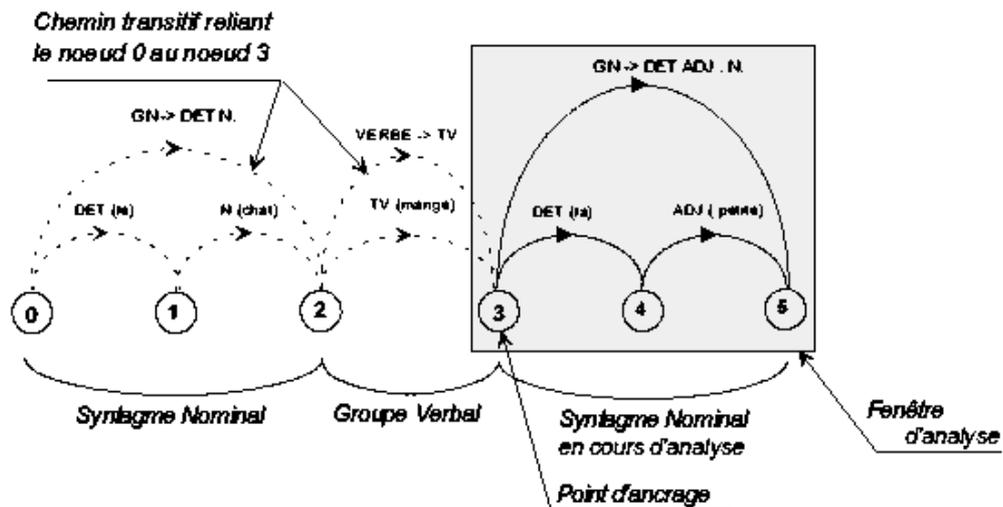


Figure 16 La segmentation (El-Guedj 1996, p. 110).

À la réception d'un nouveau mot, on applique l'algorithme du chart ascendant avec la règle d'addition de mot sur la fenêtre commençant au point d'ancrage et se terminant au dernier mot accepté. On modifie et on crée ainsi un certain nombre d'arcs. Si le mot ne peut s'intégrer dans aucun groupe de la fenêtre, alors, on déplace le point d'ancrage vers la droite au premier nœud qui satisfait aux conditions pour être un point d'ancrage.

Le principe des règles de grammaire de cette étape est de détecter tous les segments au prix d'un bruit éventuellement important. Comme résultat, l'analyse des segments génère des hypothèses de groupes. À la différence de l'analyse chomskyenne classique, elle ne va pas au-delà dans la structure de syntagme.

L'algorithme a été évalué sur un corpus de comptes rendus d'analyse médicales. Ce corpus provient du centre anticancéreux François Baclesse de Caen communiqué par le Pr. Héron. Les phrases sont essentiellement descriptives et de style varié. Le corpus total a une taille de 210.000 mots avec un vocabulaire de 10.000 formes non lemmatisées. Il nous était complètement impossible d'étudier ce corpus en entier. Nous avons retenu un ensemble tiré au hasard de 54 phrases. Ces phrases comportent 254 segments détectés manuellement.

Les règles ont été conçues manuellement en considérant les segments. La grammaire compte 66 règles. Le nombre de segments générés automatiquement est bien sûr nettement supérieur à leur nombre réel. Le taux de surdétection est d'environ 200% : sur trois segments, un seul est valide en moyenne. En revanche, le nombre de segments manqués est relativement faible : environ 6%.

3.7.8 L'élagage

L'élagage a pour but d'éliminer une partie des groupes surdétectés. En quelque sorte de nettoyer le chart. Il se fonde sur des compatibilités de bigrammes syntagmatiques. On considère ici non pas une séquence de deux mots, mais une séquence de deux groupes. Le concept de la méthode est voisin de celui de E. Brill

(1995) lorsqu'il étiquette les mots par des parties du discours. Cependant, l'action qui résulte de l'examen des groupes est soit la conservation, soit la destruction.

Pour ceci on établit le tableau des bigrammes possibles et on en définit un certain nombre d'admissibles, les autres étant inacceptables (Tableau 12). La phrase est parcourue de gauche à droite et pour un groupe donné, on vérifie qu'il existe dans le chart au moins un prédécesseur pour qu'il forme un bigramme acceptable, sinon on l'enlève du chart. On applique ensuite la réciproque dans un balayage de droite à gauche. À la fin de l'élagage, les segments qui n'ont ni prédécesseur, ni successeur, sont aussi éliminés du chart.

Bigrammes	Explications	Admissible
GN GV	Un groupe verbal peut suivre un groupe nominal	Oui
GV GV	Un groupe verbal peut suivre une groupe verbal	Non

Tableau 12 Exemples de bigrammes.

L'étape d'élagage permet de réduire les segments surdétectés à 33% du nombre total de segments dans le chart. Elle n'élimine aucun des segments corrects.

3.7.9 Les dépendances

Une fois les segments détectés, on cherche à les relier entre eux par des relations de dépendance. Ceci revient à déterminer, pour chacun des segments, quel est son régissant. La tête de la phrase, le verbe principal en général, n'ayant pas de régissant.

3.7.9.1 L'algorithme de Covington

Michael Covington (1990) a décrit une technique simple d'analyse de dépendance. Cette technique n'est pas complètement satisfaisante mais elle nous a servis de point de départ. L'algorithme utilise un ensemble de règles de dépendance semblables à celles du Tableau 3 augmentés de contraintes d'unification. Les contraintes portent notamment sur l'accord.

Les segments sont traités de gauche à droite et mis dans une liste `ListeSegmentsTraités`. La liste est initialement vide.

Une liste des régissants (`ListeRégissants`) est créée initialement vide. Elle contient les têtes à un moment donné.

On accepte les segments un par un. Pour le segment courant :

- a) Rechercher un régissant du segment courant dans `ListeSegmentsTraités` en partant des plus récents;
 - i) S'il y en a un ou plusieurs, créer la relation de dépendance sur le premier trouvé
 - ii) sinon, mettre le segment courant dans `ListeRégissants`
- b) Rechercher pour tous les segments de `ListeRégissants` les segments pouvant être subordonnés du segment courant. Lorsqu'on en trouve un, on l'enlève de `ListeRégissants`.

L'analyse a réussi quand il ne reste qu'une seule tête. On peut ajouter des contraintes supplémentaires, telles que d'éviter les croisements d'arcs ou d'enjamber une tête.

3.7.9.2 Un nouvel algorithme

L'algorithme de Covington, tel quel, n'est pas applicable aux segments stockés dans le chart. Ce chart, qui résulte des analyses précédentes, contient un surensemble des segments de la phrase et non pas des mots. Ces segments sont des hypothèses et ils ne forment pas un découpage contigu.

L'algorithme proposé reprend le principe de l'algorithme de Covington et lui apporte certaines modifications. Il utilise des règles de dépendances munies de traits d'unification et exploite la valence des groupes. Il prend en compte le nombre de segments sans tête comme critère d'évaluation avec des contraintes topologiques. Il explore les hypothèses possibles avec un mécanisme de retour en arrière.

L'algorithme gère les hypothèses de segments. Ceux-ci peuvent se recouvrir complètement ou partiellement (entrer en intersection). Pour éviter ceci, lorsqu'un segment est en cours d'analyse, on élimine les segments entièrement ou partiellement recouverts par lui.

L'algorithme procède en deux étapes en balayant le chart de gauche à droite :

1. Dans un premier temps, on recherche les relations de dépendance en suivant les contraintes de valence. Cette recherche se fait pour chaque segment ayant une valence non nulle dans le sens régissant-subordonné et dans le sens inverse ;
2. dans un deuxième temps, on recherche à établir les régissants restants pour chaque segment.

L'analyse des dépendances se fait en minimisant l'espace de recherche au départ. Pour ceci le système fonctionne en deux passes. Dans la première passe, il établit les relations certaines et dans la deuxième passe, il explore les possibilités restantes. L'analyseur de dépendance évalue la réussite de sa construction en minimisant le nombre de têtes de la phrase, une dans l'idéal. Chaque possibilité est essayée et si elle génère moins de têtes que les précédentes, elle sera retenue à sa place.

La première passe établit les dépendances certaines. Pour ceci, on tient compte de la valence des segments qu'on cherche à saturer et des relations sans ambiguïtés. L'algorithme balaye le chart et rattache les segments pour répondre successivement aux deux critères :

- un segment a une valence non vide que seul un constituant peut satisfaire dans le sens régissant-subordonné ou dans le sens inverse;
- un segment ne peut trouver qu'un et un seul régissant dans le chart. On a alors un couple régissant-subordonné.

La valence permet de guider le démarrage de l'analyse. Elle pourrait aussi contribuer à l'évaluation globale. Cependant, ce critère est moins important que celui des segments sans têtes. Des verbes transitifs, notamment, peuvent s'employer de manière intransitive. Pour cette raison, on ne l'exploite pas dans l'évaluation finale.

Dans la deuxième passe, l'algorithme établit la liste des dépendances en balayant le chart de gauche à droite et en recherchant les groupes de valence non nulle non satisfaite. Pour chacun de ces groupes, il établit la liste des groupes compatibles, dans le sens régissant-subordonné ou dans le sens inverse. Une fois que ce processus est terminé, on établit le reste des dépendances en recherchant pour chaque groupe, un régissant.

Au cours de l'analyse, les segments possibles pour une mise en relation sont classés en fonction de critères topologiques. Ces critères préfèrent considérer les groupes les plus longs d'abord quand ils partent du même mot et les liens de dépendance les plus courts.

Enfin, les arcs dépendances doivent respecter un certain nombre de contraintes, notamment :

- les arcs ne doivent pas se croiser
- un arc ne doit pas enjamber une tête
- certaines relations ne peuvent pas apparaître en double. Par exemple, un verbe ne peut pas avoir plusieurs compléments d'objet direct.

L'algorithme opère par retour en arrière du module d'établissement des dépendances vers celui de satisfaction de la valence.

3.7.9.3 Résultats

De la même manière que pour les autres modules, les règles de dépendance ont été écrites manuellement. L'algorithme a été appliqué sur le résultat produit par les modules de segmentation et d'élagage.

Sur les phrases de tests, nous avons retenu toutes les analyses plausibles d'un point de vue sémantique, même si elles ne semblent pas toujours vraisemblables compte tenu du contexte. Il y correspond des ambiguïtés d'analyse et nous avons dénombré à la main une centaine d'analyses possibles pour 53 phrases. Des compléments peuvent se rattacher soit au verbe, soit à un autre complément. Il y correspond 780 relations possibles de dépendance.

L'analyseur a considéré comme correctes les analyses conduisant au minimum de segments sans tête, en général un. Si on considère les analyses au niveau des phrases, l'analyseur a pu retrouver 80% des analyses correctes et complètes d'une phrase. Il a produit une surgénération, c'est à dire des analyses fausses, représentant 23% de ses sorties. Si on considère les relations de dépendance, l'analyseur a en retrouvé environ 90%

3.7.10 Discussion des résultats

Le système syntaxique complet (combinant les deux paradigmes de représentation) obtient des résultats encourageants et il prouve son potentiel. Nous avons pu développer une base de règles à la main qui permet d'analyser 54 phrases et nous avons pu obtenir des résultats cohérents. L'analyseur accepte et produit une analyse complète pour plus de 80% des phrases de test. Par ailleurs, nous avons pu constituer une base de règles pour l'analyseur syntagmatique seul qui permette de naviguer à la voix dans un monde virtuel.

Cependant, il se pose des questions avant de conclure complètement :

1. la première question est quantitative. Les règles ont été élaborées pour un ensemble de tests relativement petit. On a pu écrire les règles adéquates mais ceci ne prouve pas que la méthode serait valide à l'infini.
2. La seconde question est méthodologique. L'écriture des règles et leur évaluation se font sur le même corpus. On a pu écrire des règles cohérentes pour un ensemble de phrases, mais rien ne prouve encore que le résultat serait valide pour un grand nombre de phrases. On en revient à la taille du corpus.

Nous ne pouvons pas apporter de réponses, ni de solutions complètes à ces questions dans le cadre de nos moyens. Écrire des règles à la main est un travail considérable, même pour un domaine limité. Elle suppose qu'un individu ou un petit groupe puisse formaliser la plupart des phénomènes syntaxiques d'une langue. Les règles peuvent aussi se révéler difficiles à ajuster manuellement ; une nouvelle règle annulant une précédente (Church & Mercer 1993, Magerman 1994, Brill 1995). Il est vraisemblable que les résultats de notre système, tel qu'il est, baisseraient si le corpus d'analyse était plus important.

Par ailleurs, évaluer et valider un résultat à la main, même sur un petit nombre de phrases, est un travail aussi considérable. Il serait impossible de le réaliser sur un corpus un tant soit peu important. À titre de comparaison, le projet européen SPARKLE sur les grammaires peu profondes, aux moyens nettement supérieurs aux nôtres, projette une évaluation sur 200 phrases!

Une réponse complète à l'évaluation de l'analyseur demanderait à disposer de corpus annotés et parenthésés. Ces corpus permettraient aussi un apprentissage automatique de règles. Cependant, Collins (1996) ou Magerman (1994) réalisent des apprentissages de règles ou de statistiques sur des corpus de 40.000 phrases. Brill (1995) apprend ses règles sur un million de mots. Il n'existe rien d'équivalent pour le français à l'heure actuelle. Ceci est incontournable pour le développement systématique de règles d'analyse syntaxique ainsi que pour une évaluation rigoureuse.

3.8 Conclusion et perspectives

Dans ce chapitre, nous avons décrit un système d'analyse syntaxique se fondant sur des règles syntagmatiques et des règles de dépendance. La partie syntagmatique de l'analyseur peut être rendue autonome. Elle a été intégrée dans un agent conversationnel que nous détaillons dans le prochain chapitre.

Comme perspective, le système développé a démontré sa cohérence en obtenant des résultats positifs sur la base de test. L'analyseur est aussi en phase avec la démarche actuelle d'autres équipes de recherches :

- pour ce qui est de la première couche d'analyse qui détecte les groupes, l'analyseur utilise une grammaire syntagmatique. Ceci est semblable à la méthode de beaucoup d'équipes de recherches désormais. On peut sans doute dater cette orientation des résultats du système FASTUS lors des concours MUC et de la

justification éloquente d'Appelt et al. (1993). Ceci a eu comme conséquence un regain d'intérêt considérable pour l'analyse syntaxique de surface.

- Pour ce qui est de l'architecture globale, l'analyseur fonctionne suivant un principe de couches. Il est très voisin dans sa structure de celui de Collins (1996). Comme lui, notre analyseur cherche ensuite à établir des relations de dépendance entre des hypothèses de segments contenues dans un chart. Le système de Collins est cependant fondé sur des méthodes d'évaluation stochastiques alors que notre système est symbolique.

Les méthodes stochastiques ont été les premières mises en œuvre pour réaliser une analyse syntaxique obtenant de bons résultats sur des grands corpus. Elles s'inspiraient et faisaient suite aux travaux en reconnaissance de parole. Elles ont principalement été appliquées à l'étiquetage par les parties du discours.

Les méthodes symboliques connaissent une certaine popularité depuis les années 1990 et l'analyseur d'Eric Brill. Ces méthodes obtiennent des résultats comparables en termes de précision aux méthodes stochastiques. Elles peuvent dans certains cas se transformer dans des automates d'états finis et elles sont alors dix fois plus rapides (Roche & Schabes 1995).

Une partie des règles du système que nous avons présenté pourraient s'apprendre automatiquement à partir d'un corpus annoté et parenthésé (à condition d'en avoir un disponible). Cet apprentissage permettrait s'affranchir de l'écriture de règles à la main. Ceci offre la perspective de réaliser un analyseur syntaxique symbolique du français à large couverture.

4.
Agents conversationnels dans un
monde virtuel

4.1 Présentation

Nous avons abordé l'interaction linguistique dans le cadre d'un prototype de dialogue oral dans un domaine médical (Nugues et al. 1991; Nugues et al. 1992; Nugues et al. 1993(a); Nugues et al. 1993(b); Cazenave et al. 1994). Ce prototype – *DictaMed* – permettait à un utilisateur de dicter des comptes rendus médicaux dans un domaine restreint – l'analyse de Holter tensionnels – puis d'interroger le système sur le contenu de ces comptes rendus.

DictaMed utilisait un dispositif de reconnaissance vocale du commerce ainsi qu'un autre de synthèse. Il comportait des modules syntaxique et sémantique qui analysaient la dictée et il générait automatiquement le document correspondant. *DictaMed* se composait d'un noyau-objet pour représenter les concepts linguistiques et médicaux; d'un analyseur syntaxique; d'un analyseur sémantique gérant les entités médicales.

En comparaison avec d'autres systèmes de dictée, *DictaMed*, avait la capacité d'interagir oralement. L'analyseur syntaxique, notamment, acceptait les formes interrogatives et, grâce à l'analyseur sémantique, *DictaMed* pouvait répondre à certaines questions concernant des entités précédemment décrites. Pour ceci, l'analyseur sémantique déterminait les objets et les attributs sur lesquels portait la question et recherchait les valeurs de ces attributs. L'interface de dialogue pilotait alors la synthèse des messages audios de réponse à la question. Cette interface pouvait aussi avertir l'utilisateur d'erreurs.

Les concepts de *DictaMed* ont servi de base à la construction d'un agent conversationnel plus élaboré – *Ulysse* – permettant de naviguer à la voix dans un monde virtuel (Diebel 1994; Revolta 1995; Bersot et al. 1996; Bersot 1996; Nugues et al. 1996; Godéreaux et al. 1996(a); Godéreaux et al. 1996(b); Godéreaux 1997; Godéreaux et al. 1997).

Nous avons conçu et implanté *Ulysse* dans le cadre du projet européen COST-14 (Cotech 1995). COST-14 a eu comme objectif l'étude des environnements virtuels et augmentés pour le travail coopératif à l'aide de l'ordinateur – *Computer Supported Cooperative Work* en anglais – (Grudin 1994).

4.2 Le travail coopératif

Les systèmes de travail coopératif fournissent les outils à un groupe de personnes dispersées géographiquement pour travailler ensemble. Ces systèmes se fondent sur l'utilisation d'un réseau et leur diffusion est maintenant importante dans certaines organisations et entreprises. Le travail coopératif est par ailleurs porté par la croissance d'Internet. L'étude de COST-14 a notamment examiné comment intégrer

des communications synchrones rapides telles que la visioconférence dans ce type d'environnements.

Les environnements de travail coopératif tentent de rappeler certains objets d'un bureau en incorporant des icônes, des plans, des fonds d'écran symboliques. De plus en plus, ces environnements font appel à des techniques de réalité virtuelle³. Grâce à elles, on peut reconstituer des bureaux ou des espaces de travail avec leurs accessoires familiers. Ces espaces virtuels sont partageables entre plusieurs utilisateurs grâce à des techniques d'informatique répartie. Autour de ces techniques se construisent des normes telles que VRML et Java.

La visioconférence incorporée dans un environnement virtuel permet de recréer des salles de réunion, de visualiser et de rendre intelligible des scènes parfois complexes (Figure 17, Figure 18). Dans les mondes virtuels, les utilisateurs sont figurés par des « clones » plus ou moins réalistes. On nomme ces figures des représentations, des incarnations ou encore des avatars⁴ – la terminologie n'est pas encore bien établie. Les pionniers des techniques de travail en groupe dans le domaine du bureau sont les laboratoires Xerox au PARC (Robertson et al. 1993).

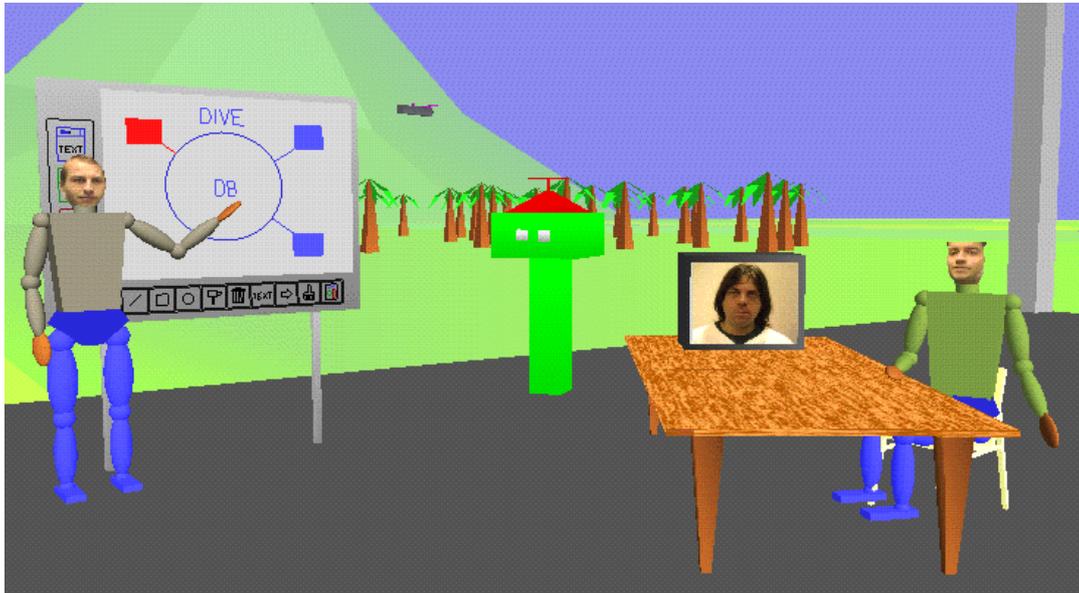
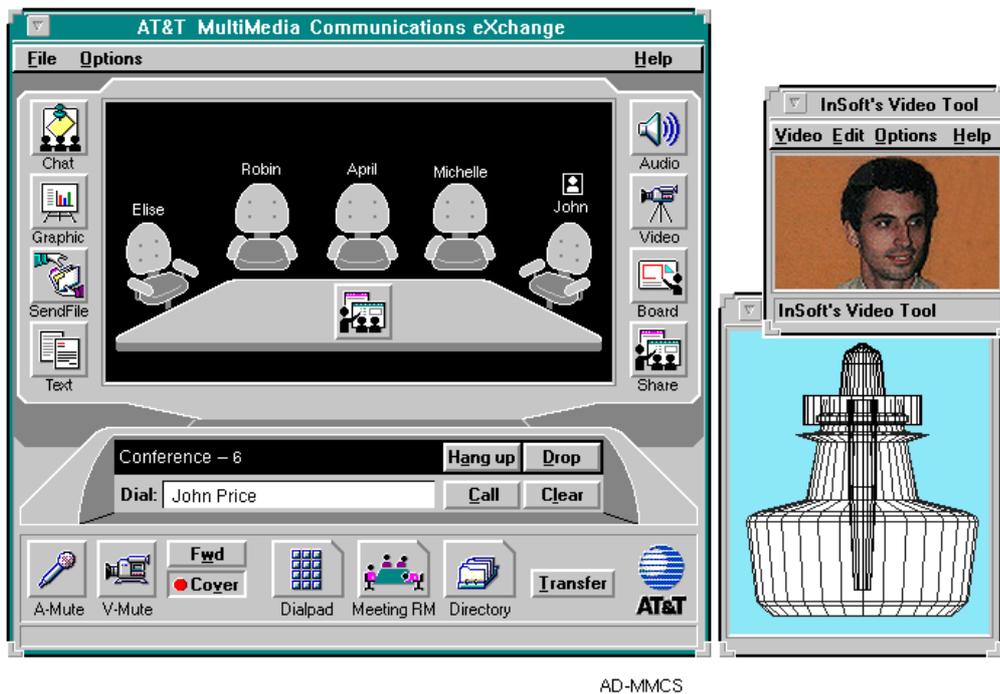


Figure 17 Visioconférence dans un bureau virtuel (Benford et al. 1995).

³ L'expression réalité virtuelle est parfois contestée – on la qualifie d'oxymore – et certains préfèrent parler de simulation interactive dans un monde tridimensionnel.

⁴ Un avatar est une incarnation de Vishnou dans la religion hindoue.



AD-MMCS

Figure 18 Visioconférence 2-D (ATT).

4.2.1 Pourquoi un environnement virtuel et comment le piloter

Les avantages de la réalité virtuelle résident dans l'image et l'interactivité. Les images grâce à leur puissance cognitive facilitent la compréhension d'une situation, la communication d'une idée. Dans le travail coopératif, on utilise des métaphores d'objets courants et d'environnements de travail pour aider l'utilisateur à accéder au sens immédiat d'une scène. Ces métaphores peuvent être bidimensionnelles (Figure 17) ou bien spatiales – tridimensionnelles – lorsqu'on utilise la réalité virtuelle (Figure 18).

L'interactivité élimine l'impression de passivité éventuelle d'un utilisateur. Elle lui donne une plus grande liberté dans l'exploration et la navigation. Et en écho au proverbe chinois :

J'entends et j'oublie, je vois et je me souviens, je fais et je comprends.

de tels environnements permettent de voir et de faire. La supériorité cognitive de ce type d'environnements semble aller de soi et pourtant Philippe Quéau (1996) affirme que par une sorte de dogmatisme, certains chercheurs ont ignoré la valeur de l'image.

Dans les environnements virtuels, l'interaction se réalise presque toujours avec des systèmes de pointage, tels que des souris, des boules tridimensionnelles, des manettes, etc (Figure 19). Ces dispositifs permettent la navigation par des mouvements dans des plans horizontaux ou verticaux et des rotations (Figure 20). De tels dispositifs permettent aussi de pointer un objet particulier et de se téléporter

vers lui. Finalement, les pointeurs permettent de piloter les objets du monde virtuel : de les manipuler, de les activer, etc.

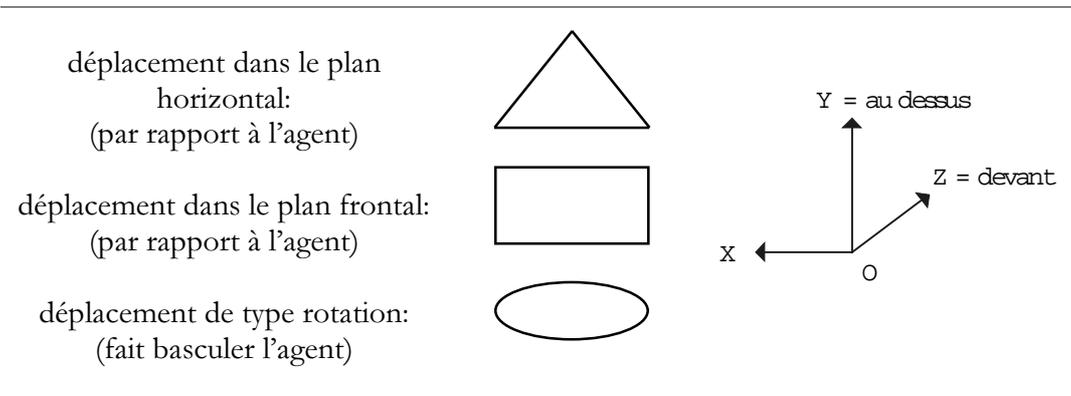


Figure 19 Les pointeurs de DIVE pour le « véhicule » souris.

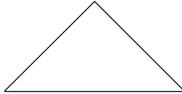
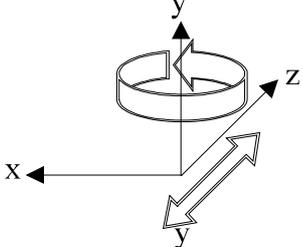
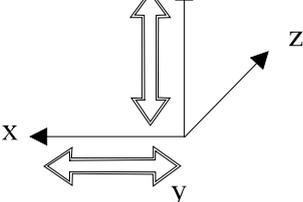
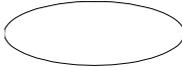
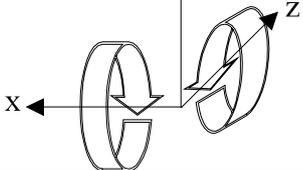
Type de déplacements	Icônes	Représentation
déplacement dans le plan horizontal : un axe de translation, un axe de rotation		
déplacement dans le plan frontal : deux axes de translation		
deux axes de rotation		

Figure 20 Les déplacements associés.

L'interaction dans les mondes virtuels est délicate. C'est en quelque sorte la contrepartie de la meilleure compréhension. Les systèmes de pointage classiques permettent une désignation rapide et précise des objets à condition que les objets en cause soient dans le champ de vision. Le déplacement de l'utilisateur – la navigation – est assez facile si le mouvement est rectiligne ou géométriquement pur. Les visites de mondes virtuels, dans les séquences de démonstration ou les films grand public, sont d'ailleurs remplies de vols planés. En revanche, on y trouve peu de mouvements

« naturels », tels que des rotations, des coups d'œil à gauche ou à droite. Ces mouvements sont difficiles à réaliser avec les pointeurs classiques. Ceci sans parler de gestes articulés tels que se déplacer en fixant un objet du regard.

Nous avons constaté dans le cadre de nos essais et lors de démonstrations que la navigation représente la difficulté centrale de l'interaction dans environnements virtuels de visioconférence (Godéreaux et al. 1996). Cette navigation est même pratiquement impossible pour les utilisateurs débutants. Ils y attrapent facilement le mal de mer, leur « corps » virtuel sens dessus dessous en quelques secondes.

4.2.2 Interfaces linguistiques dans un environnement virtuel

Les interfaces à des mondes virtuels ou à des environnements de simulation sont nombreuses. Elles mettent en jeu essentiellement des dispositifs d'affichage et de pointage manuel. Certains dispositifs ont des noms qui relèvent de la science fiction : visiocasque, exosquelette, capteur haptique, etc. Cependant, les systèmes permettant une interaction orale sont plus rares. Les plus notables sont *Diverse* (Karlgrén et al. 1995), *Persona* (Ball et al. 1997), *Nautilus* (Everett et al. 1995), *TRAINS* (Allen et al. 1995) et *Put-That-There* (Bolt 1980). Certains de ces systèmes ont une voie d'entrée exclusivement linguistique : ils n'acceptent que les requêtes orales; même si les réponses de l'agent sont à la fois auditives et visuelles. C'est le cas de *TRAINS*, *Nautilus* et *Persona*. D'autres sont multimodaux et fusionnent la voix et le geste de l'utilisateur. C'est le cas de *Diverse* et de *Put-That-There*.

Parmi ces systèmes, *Persona* et *TRAINS* ont un domaine d'application clairement défini faisant appel à des capacités délimitées : *Persona* est un agent conçu pour rechercher des morceaux de musique dans un juke-box et les jouer. L'agent s'anime sous les traits d'un perroquet et dialogue avec l'utilisateur. On peut l'interroger sur des styles de musique, des compositeurs. *TRAINS* est un planificateur conversationnel qui aide l'utilisateur à assembler des trains et choisir des routes pour transporter des oranges d'un entrepôt à une usine de pressage. *TRAINS* affiche les routes sur un écran en même temps qu'a lieu le dialogue. Le planificateur a une connaissance sur les emplacements des locomotives et des voitures, ainsi que sur la configuration d'un convoi. Cependant ses capacités de raisonnement sont limitées pour forcer à engager le dialogue et l'interaction.

L'interaction, le dialogue et surtout le raisonnement sous-jacent dans *Diverse* et *Nautilus* ont des contours plus vagues : le génie de *Diverse* manipule les objets d'un monde virtuel. On peut lui demander de repeindre une maison. Il peut aussi naviguer – déplacer la représentation de l'utilisateur – de façon limitée. On ne sait cependant pas quelles sont les limites des manipulations. Il est peu vraisemblable cependant que *Diverse* accepte des requêtes telles que : *mets le toit de la maison à l'envers sur la vache*. *Nautilus* est une interface vocale à un simulateur de champ de bataille aérien. On peut l'interroger sur la configuration des avions alliés et ennemis. De même, que pour le projet précédent, on ne sait pas les limites de sa connaissance et ni ses capacités. Le lexique étant de 300 mots, on peut imaginer que les questions posées doivent rester simples. *Put-That-There* avait des capacités plus limitées, mais il est beaucoup plus ancien.

Les interfaces linguistiques à des mondes virtuels utilisent aussi des techniques voisines de certaines employées pour le pilotage des robots à la voix. Cependant le

pilotage de robots n'intègre pas l'interactivité. Crangle & Suppes (1994) donnent une synthèse de ce domaine.

4.3 Un corpus de dialogues

4.3.1 Les conditions de l'expérience

Pour préciser l'intérêt d'une interface linguistique, nous avons rassemblé un corpus de dialogues où quatre utilisateurs ont interagi avec un monde semblable à celui de la Figure 21 (Godéreaux et al. 1996, Godéreaux 1997). Chaque dialogue a impliqué deux participants : l'utilisateur et un opérateur. L'opérateur jouait le rôle de l'agent en agissant sur le monde virtuel en fonction des ordres de l'utilisateur. Les dialogues ont mis en jeu deux utilisateurs débutants, capables de manipuler un traitement de texte, et deux utilisateurs informaticiens, sans connaissance particulière des mondes virtuels.

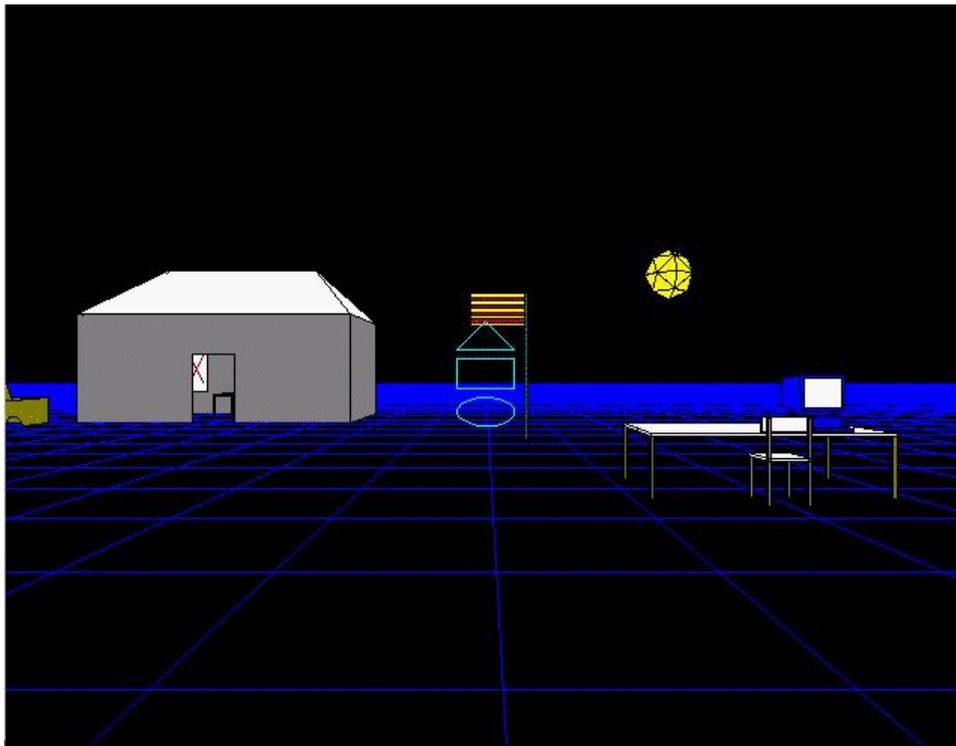


Figure 21 Le monde virtuel Ithaque.

Nous avons rassemblé les corpus en utilisant l'environnement virtuel interactif distribué DIVE (Andersson et al. 1994) de l'Institut suédois d'informatique (SICS). DIVE permet de construire des mondes virtuels où les utilisateurs peuvent se

connecter d'un endroit distant, se déplacer et rencontrer d'autres participants. Les participants partagent le même modèle géométrique avec un point de vue différent. Les modifications apportées au monde provenant des interactions des utilisateurs sont répliquées sur les autres sites participants pour maintenir le monde cohérent.

Nous avons conçu deux scénarios. Dans le premier, les utilisateurs débutants devaient découvrir le monde et dans le second, les utilisateurs plus expérimentés devaient trouver un trésor caché dans le monde. Dans le cadre de cette expérimentation, l'utilisateur était seul dans le monde – il n'y avait pas d'autres participants connectés.

4.3.2 L'intérêt d'une interaction orale

En comparant les interactions vocales et à la souris, nous nous sommes rendu compte que la navigation à la souris était une difficulté majeure. Dans des sessions ultérieures, nous avons même constaté que la plupart des utilisateurs débutants étaient incapables de faire le tour d'un objet. De manière plus précise, si la navigation est relativement aisée dans un plan donné, il est beaucoup plus difficile de s'aligner ou d'effectuer un mouvement circulaire. De plus, il est impossible de regarder un endroit particulier pendant un déplacement avec la seule souris. Ceci rend certains mouvements malaisés, par exemple quand un utilisateur fait le tour d'un objet.

À l'opposé, beaucoup de mouvements sont faciles à formuler verbalement (Tableau 13, Tableau 14). La coordination de l'entrée vocale et d'un pointeur permet à un utilisateur d'aller vagabonder dans tous les recoins d'un monde virtuel.

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

Tableau 13 Un dialogue avec alignement.

A	<i>nous sommes connectés au monde robot.</i>
U8	<i>tourne sur toi même.</i>
A	<i>vers la droite ou vers la gauche?</i>
U9	<i>vers la droite.</i>
A	<i>voilà.</i>
U10	<i>prends de la hauteur.</i>
U11	<i>arrête de monter.</i>
U12	<i>monte.</i>
A	<i>oui.</i>
U13	<i>stop.</i>

Tableau 14 Extrait du corpus du dialogue.

En conclusion du dépouillement de ce corpus (Godéreaux et al. 1996), nous nous sommes rendu compte que les interfaces linguistiques rendent les environnements virtuels plus commodes, plus aisés à manipuler :

- au-delà de l'image, elles immergent l'utilisateur dans le son et la parole;
- elles facilitent la navigation;
- elles ajoutent un canal d'interaction.

Ces trois critères – immersion, navigation et interaction – correspondent selon Philippe Quéau (1993) aux trois axes d'évaluation des systèmes de réalité virtuelle.

4.4 Intégrer le dialogue oral dans un monde virtuel

L'interaction orale entre une personne et une machine nécessite plusieurs modules ou dispositifs linguistiques relativement génériques tels que des systèmes de reconnaissance et de synthèse de parole, des analyseurs syntaxique, sémantique et un gestionnaire de dialogue (Allen 1995). Dans un environnement virtuel, la parole n'est qu'un mode d'interaction parmi d'autres – dans certaines circonstances un mode mineur – et on doit apporter certaines adaptations aux architectures de dialogue classiques.

Une différence importante est l'intégration d'un système de pointage avec la parole. Il ne semble pas désirable d'essayer de substituer complètement les systèmes de pointage classiques : il est souvent plus facile de désigner du doigt un objet que de le décrire de façon verbale. Il y a d'autres situations où l'on préfère « le dire » que de « le faire ». Pour cette raison, l'interface doit laisser à l'utilisateur le choix – l'initiative – du moyen d'interaction qu'il veut utiliser. L'interface doit fusionner les deux modes d'interaction – la désignation gestuelle et la parole – pour que l'un puisse compléter l'autre. Ceci implique notamment des moyens pour résoudre les références déictiques issues de la parole ou du geste et les coordonner entre elles.

La résolution des références oblige *in fine* à raisonner sur les objets de la scène et leur forme. La compréhension d'une situation implique de faire correspondre des modèles géométriques d'objets à des descriptions linguistiques de l'espace.

Enfin, l'architecture d'une interface orale doit comprendre un gestionnaire d'actions qui agira au nom de l'utilisateur dans le monde virtuel. Le gestionnaire devra planifier des mouvements suffisamment « humains » et prévisibles, lorsqu'il s'agit d'une navigation ou d'une manipulation, pour que l'utilisateur se sente à l'aise dans son « incarnation » et avec les objets qui l'entourent.

4.5 Ulysse

4.5.1 L'architecture

Ulysse prend la forme d'un agent conversationnel qui s'incorpore dans l'incarnation de chaque utilisateur. Les capacités d'Ulysse sont relativement

spécifiques et ne concernent que la navigation. Ulysse assiste l'utilisateur à l'intérieur du monde en répondant positivement aux commandes de mouvement. Il agit en conséquence et transporte l'utilisateur en son nom à l'intérieur de l'environnement virtuel.

Dans d'autres projets tels que celui de J. Karlgren et al. (1995) et de S. Everett et al. (1995), des capacités multiples sont implantées qui permettent à l'utilisateur de « parler » au monde. Les agents correspondants agissent en fonction du contexte et normalement naviguent (déplacent l'incarnation de l'utilisateur), manipulent des objets virtuels ou répondent à des questions. Ce type de conception rend peut-être les agents moins compréhensibles pour l'utilisateur. Il cerne moins bien les capacités de raisonnement de l'agent. L'utilisateur pourrait s'étonner à un moment donné : Qu'est-ce que je peux demander à mon agent?

La structure globale d'Ulysse s'inspire du prototype DictaMed réalisé auparavant. Elle est semblable à celles d'autres systèmes de dialogue interactif (Allen et al. 1994). Elle comprend des dispositifs de reconnaissance et de synthèse vocale, un analyseur syntaxique, des modules sémantique et de dialogue (Figure 22). L'architecture d'Ulysse est aussi déterminée par le raisonneur du domaine, le planificateur et le gestionnaire d'actions. Ces deux derniers modules sont propres à chaque modèle d'utilisateur et dépendent des capacités qu'on lui aura attribuées. Une incarnation humaine n'aura pas le même mouvement qu'une incarnation de voiture.

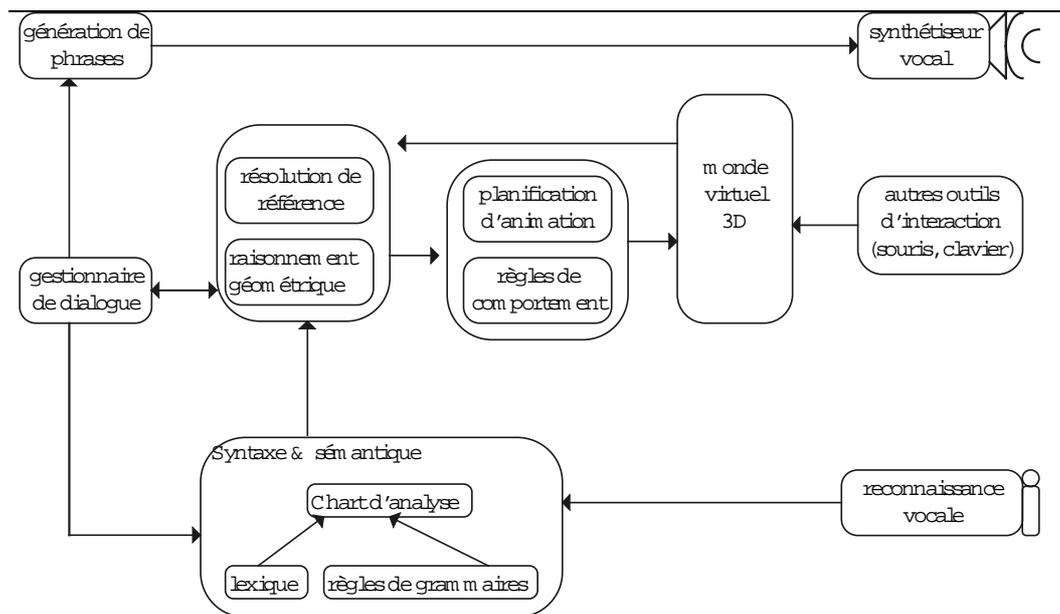


Figure 22 L'architecture du système Ulysse.

À la différence de TRAINS (Allen et al. 1994, p. 18), la désindexation est étroitement liée au dialogue et aux capacités de raisonnement. La compréhension des commandes de navigation nécessite de résoudre les nombreuses références déictiques qui apparaissent dans la conversation et de raisonner sur la géométrie du monde. L'architecture d'Ulysse est complétée par un résolveur de références qui se coordonne avec les gestes de l'utilisateur et ses énoncés, et un raisonneur géométrique pour comprendre le monde. L'utilisateur peut ainsi nommer et pointer

des objets. Enfin, la navigation est réalisée par un planificateur et un gestionnaire d'actions qui amène l'utilisateur dans un mouvement relativement continu où il veut aller.

4.5.2 La reconnaissance vocale

La précision des systèmes de reconnaissance vocale s'est améliorée de façon considérable ces dernières années. Ceci est notamment dû à la disponibilité de très grands corpus de texte et de parole et à l'adoption généralisée de méthodes statistiques (Church & Mercer 1993).

La reconnaissance peut s'effectuer sur de la parole continue – les mots sont prononcés normalement – ou bien discrète. Le locuteur doit alors marquer une pause entre les mots. Les meilleurs systèmes actuels sont programmables et peuvent reconnaître les deux types de parole : continue ou bien discrète. La reconnaissance continue demande qu'on ajoute une grammaire syntagmatique. Elle porte alors sur un nombre de mots limités. Dans le cas discret, les systèmes utilisent des modèles trigrammes qu'ils ont en mémoire et leur vocabulaire est beaucoup plus important.

Nous avons travaillé avec plusieurs systèmes commerciaux de reconnaissance et en particulier le système VoiceType d'IBM. VoiceType opère sur des mots isolés, et il se destine en premier lieu à la dictée de rapports. Nous l'avons choisi parce qu'il traite le français ainsi que d'autres langues européennes avec un vocabulaire allant jusqu'à 30.000 mots.

VoiceType ne fonctionne que sur un PC. Dans le cadre d'Ulysse, les mots reconnus sont transmis, phrase par phrase, par le réseau local vers la station de travail où s'exécutent l'agent et le monde virtuel de DIVE. On déclenche la transmission en pressant un bouton. Les mots sont ensuite récupérés par l'analyseur syntaxique.

4.5.3 L'analyse syntaxique

L'analyseur syntaxique d'Ulysse est celui décrit au chapitre précédent. Il est fondé sur les charts et adopte l'algorithme ascendant classique avec un formalisme double. Il peut opérer avec des règles syntagmatiques classiques ainsi qu'avec des règles de dépendance. Pour Ulysse, nous n'avons utilisé que la partie syntagmatique. L'analyseur et le système de reconnaissance fonctionnent sur deux machines différentes. Ils sont reliés entre eux par un réseau. L'analyseur accepte de gauche à droite les mots reconnus.

Une grammaire de constituants a été utilisée pour coder le lexique – 350 mots distincts – et des règles syntagmatiques qui acceptent l'ensemble des 400 énoncés du corpus (Godéreaux et al. 1994; Godéreaux et al. 1996). Le lexique utilise des étiquettes pour les parties du discours qui sont une variante des catégories de MULTEXT (Ide & Véronis 1995). Nous n'avons retenu que les catégories pertinentes pour le français.

Les règles syntagmatiques récrivent les énoncés en utilisant des contraintes d'unification et des catégories non terminales telles que les groupes nominaux, les groupes verbaux, les groupes prépositionnels, les groupes déterminants, les groupes adverbiaux, les groupes d'adjectifs, etc. Les règles ont été adaptées pour accepter des mots manquants ou inconnus. Elles comprennent un grand nombre de locutions

prépositionnelles, adverbiales et démonstratives qui sont fréquentes dans la langue parlée.

Les énoncés se décomposent en quatre types principaux de propositions : les ordres, les questions, les affirmations, les subordonnées et les énoncés sans verbe. L'utilisateur segmente ses énoncés phrase par phrase après la reconnaissance vocale en pressant un bouton. L'analyse syntaxique résulte dans un ou plusieurs arbres d'analyse. Il y en a jusqu'à huit dans notre corpus; ils reflètent l'ambiguïté syntaxique ou sémantique de l'énoncé.

4.5.4 Les actes de dialogue et le traitement sémantique

Les arbres d'analyse des énoncés sont d'abord mis en correspondance avec des actes de conversation qui représentent principalement des commandes de navigation telles que *va dans la maison*. Les autres actes de conversation qui sont identifiés dans Ulysse sont :

- les clarifications déictiques telles que *celle-ci*
- les modifications de mouvement telles que *plus vite*
- les répétitions de mouvement telles que *encore*

Le module d'interprétation sémantique ne considère que les commandes de navigation. Il partage l'énoncé en propositions et annote les constituants de l'arbre qu'il extrait du chart avec des fonctions syntaxiques. Ces fonctions sont classiques et correspondent aux sujet, verbe, compléments d'objet, compléments circonstanciels. Cette étape rattache aussi les adverbes au mot qu'ils modifient : un verbe ou un autre adverbe.

Nous avons étiqueté certains groupes avec des catégories sémantiques. Les compléments circonstanciels sont annotés avec leurs catégories ontologiques classiques. L'annotation sémantique des verbes est en relation avec le mouvement – la navigation – que demande l'utilisateur et à la description de l'espace.

En considérant le corpus et en dépouillant systématiquement une source lexicale verbale (Bescherelle 1980), nous avons divisé les verbes en six catégories sémantiques principales (Tableau 15). Les deux premières catégories correspondent à des changements de lieu et les deux secondes à des changements de postures (Sablayrolles 1995).

Catégories	Verbes	Commentaires
Aller	<i>aller, avancer, entrer, monter, sortir, etc.</i>	Ces verbes correspondent à un changement de lieu avec une rotation possible de l'incarnation
Retourner	<i>revenir, retourner, etc.</i>	Cette catégorie est semblable à la précédente mais elle ne tient pas compte de la visibilité de l'objet dans le raisonnement ultérieur
Tourner	<i>tourner, obliquer, pivoter, etc.</i>	Ces verbes correspondent à une rotation du corps tout entier.
Regarder	<i>regarder</i>	Ces verbes correspondent à une rotation de

		la tête
Arrêter	<i>arrêter, stopper, etc.</i>	On arrête l'action en cours
Continuer	<i>continuer, reprendre, etc.</i>	On reprend l'action précédente

Tableau 15 Les catégories de verbes.

L'affectation d'étiquettes sémantiques est parfois ambiguë. Comparons *retourne-toi* qui appartient à la 3^e catégorie et *retourne dans la maison qui correspond* à la 2^e. Nous avons diminué les ambiguïtés en utilisant les sous catégories syntaxiques du verbe : transitif, intransitif ou pronominal, et en les imposant comme des contraintes d'unification.

Le résultat de cette étape est que chaque phrase est transformée en une liste avec autant d'éléments qu'il y a de propositions. Chaque proposition est mise en correspondance avec une structure dont les membres sont le sujet, le groupe verbal et une liste de compléments. Chaque complément étant annoté avec une étiquette sémantique : temps, lieu, manière, etc. Les groupes verbaux sont aussi annotés avec une étiquette de mouvement et incorporent leurs pronoms clitiques et d'éventuels adverbes. Ulysse fait correspondre à la même commande des énonciations différentes telles que :

Avance
Je veux avancer
Peux-tu avancer?
Je veux que tu avances

Quand les énoncés consistent en plusieurs propositions, celles-ci sont concaténées et éventuellement réorganisées suivant leurs « connecteurs ». Ces connecteurs sont associés à des opérateurs de liste tels que ajoute, détruit, remplace ou insère. Les connecteurs peuvent être des adverbes, des conjonctions, ou des formes syntaxiques. Par exemple, l'adverbe de négation *non* dans la phrase *Monte non descends* résulte dans le remplacement du premier verbe. L'adverbe *puis* dans la phrase *Monte sur la maison puis va devant l'ordinateur* résulte dans l'ajout de la seconde action. Le participe présent *passant* dans la phrase *Va vers la maison en passant devant le drapeau* a pour résultat l'insertion du dernier mouvement avant le premier.

La forme logique est ensuite traitée pour la relier à une séquence d'actions de base. Suivant le type du verbe, une proposition peut être transformée en une ou plusieurs actions de base (jusqu'à quatre pour *contourner*). Par exemple :

Monte sur le drapeau correspond à :

1. aller sur <objet>

Retourne dans la maison correspond à :

1. tourner de 180°
2. aller dans <objet>

Pour ceci les verbes complexes sont définis en termes de primitives verbales plus simples. Par exemple *contourner* se définit comme aller vers l'objet, aller à droite de l'objet puis aller de l'autre côté de l'objet.

L'agent est articulé et son regard peut se diriger dans un sens différent de l'orientation de son corps ou de son déplacement. Le corps reste fixe et la tête tourne dans l'ordre *regarde à gauche*. Dans un ordre tel que *fais le tour de la maison*, l'agent garde un œil sur l'objet qu'il a mentionné pendant qu'il se déplace

Dans le dernier cas, on doit piloter simultanément un changement de lieu et un changement de posture. Pour ne pas perdre de généralité, le module sémantique complète chaque élément de la suite de formes logiques par les changements de lieu ou les changements de postures non explicites, même s'ils sont inexistantes. Il en résulte une séquence finale dont chaque élément contient les triplets :

- aller <préposition> <objet ou direction>
- se tourner (direction du corps) <préposition> <objet ou direction>
- regarder <préposition> <objet ou direction>

4.5.5 Le raisonnement géométrique et la résolution des références

Le module de résolution de références désindexe la séquence de prédicats d'actions qui résultent de l'interprétation sémantique. Les références d'objets sont omniprésentes dans le corpus et dans toutes les sessions d'interaction que nous avons réalisées par la suite. Elles comprennent :

- des éléments ou des objets spécifiques tels que *va devant l'ordinateur*, *rentre dans la voiture à gauche de la maison*,
- des pluriels: *dirige-toi vers les voitures*,
- des choix multiples: *va dans la maison* (avec plusieurs maisons) et
- des énoncés déictiques tels que *va ici*.

Pour résoudre ces références, l'agent doit prendre en compte l'état de la base de données du monde, la position de l'utilisateur dans le monde, ainsi que l'historique de l'interaction.

L'association d'un nom à un objet est parfois délicate. Les utilisateurs peuvent utiliser des mots différents pour désigner la même chose. Les bases de données géométriques peuvent aussi considérer certains objets comme des composés ou des hiérarchies alors qu'ils forment des entités uniques dans l'esprit de l'utilisateur. Pas exemple, une maison peut se représenter comme un objet unique, comme des fenêtres, des portes, des murs, etc. ou comme un ensemble de lignes polygonales. Des systèmes tels que WordNet (Miller 1995) tentent de formaliser ces structures.

Avec Ulysse, nous avons traité le problème de la dénomination en associant soigneusement un nom aux entités de la base de données. Nous avons structuré cette base pour conserver les relations les plus cohérentes entre les noms et les entités du monde en fonction de notre corpus. Nous avons aussi muni chaque entité d'axes de références dont l'origine est le centre de gravité de l'objet.

Quand un nom d'objet apparaît dans un énoncé, Ulysse le référence en extrayant, de la base de données géométrique, une liste d'entités compatibles du point de vue du sens. Il considère le groupe nominal en entier et applique incrémentalement des

contraintes en fonction des noms, des adjectifs et des groupes prépositionnels. Les groupes prépositionnels sont analysés de droite à gauche. Par exemple

va vers le cube rouge à droite de la maison

entraînera la recherche d'une maison, puis un cube rouge à droite de cette maison.

Quand plusieurs objets sont candidats, le module résout l'ambiguïté par un algorithme de saillance semblable à celui de Huls et al. (1995). Deux critères sont pris en compte suivant le type du verbe :

- la visibilité des objets du point de vue de l'utilisateur
- un focus qui reflète l'historique des interactions de l'objet

La visibilité d'un objet résulte de l'intersection du cône de visibilité de l'utilisateur avec la base de données du monde (Figure 23). Le focus est un entier attaché à chaque objet. Ulysse modifie le focus d'un objet à chaque fois que l'utilisateur interagit avec lui – le mentionne oralement ou le pointe avec la souris – et il lui affecte comme valeur le maximum de tous les focus plus un. Ce focus est similaire à celui décrit par J. Karlgren et al. (1995).

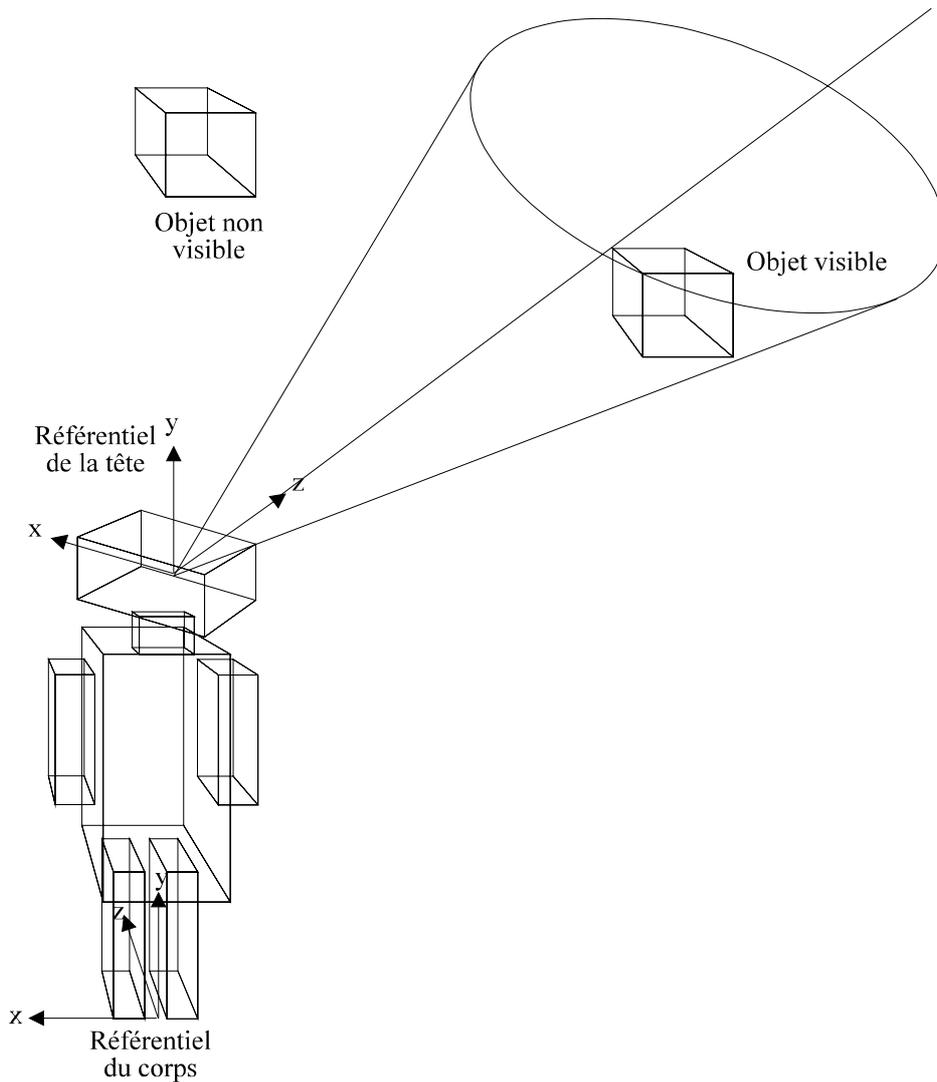


Figure 23 Le cône de visibilité d'un utilisateur.

En fonction de la catégorie du verbe et de la préposition, Ulysse applique un ou deux critères pour éliminer les ambiguïtés sur les objets. Pour les couples (verbe, préposition), tels que *aller vers*, *avancer devant*, on fait l'hypothèse que l'objet est visible et le module de résolution des références supprimera les objets non visibles. Puis si l'objet n'est pas unique, il retiendra le plus grand focus. Il retiendra l'objet ayant le plus grand focus pour les autres couples tels que *retourner vers*, *revenir devant*.

Bien que le mécanisme de saillance semble simple, nous parvenons à la même conclusion que Carla Huls et al. (1995) : il est efficace et semble donner d'aussi bons résultats que celui de Grosz et Sidner (1986).

Une fois les objets référencés, le module calcule une suite de positions géométriques tridimensionnelles correspondant à la séquence des formes logiques. Pour ceci, il tient compte des triplets (verbe, préposition, objet). Les verbes issus du module sémantique correspondent aux trois catégories : *aller*, *regarder* et *se tourner*.

L'algorithme utilisé pour *regarder et se tourner* correspond simplement à orienter le corps ou le regard dans une direction ou vers un objet.

Pour ce qui concerne les verbes de déplacement, nous avons classé les objets suivant différentes formes. Nous avons distingué ceux qui ont un « devant » et un « derrière », des objets qui ne sont pas orientés. Nous avons ainsi donné une orientation aux axes de référence de chaque objet s'il pouvait en avoir une. Nous avons aussi tenu compte de leur taille. La forme générale et le centre de gravité des objets permettent de calculer une sorte de distance acceptable pour positionner un utilisateur relativement à un objet quand il veut se déplacer vers lui : proche pour des petits objets, plus lointaine pour un objet plus grand. Les centres de gravité permettent aussi d'approximer un groupe d'objets – qui apparaît dans un pluriel – à une entité unique.

Les prépositions sont celles relevées dans le corpus. Nous les avons classés en huit groupes : *devant, derrière, vers, sur, sous, dans, hors de, à gauche (à droite)*. Pour chaque type d'objet et pour chaque type de préposition, nous avons associé un déplacement caractéristique.

Ces déplacements permettent de transformer la liste de formes logiques en un tableau de positions tridimensionnelles où chaque élément décrit :

- la direction où l'agent dirige son regard;
- la direction où l'agent dirige son corps;
- l'endroit où l'agent se dirige.

4.5.6 La gestion du dialogue

Le module de dialogue règle les tours de parole et le séquençage des modules d'Ulysse. Ceci correspond à la saisie des énoncés, à leur traitement et à leur exécution. Le module de dialogue gère les ambiguïtés syntaxiques en fournissant séquentiellement les arbres d'analyse syntaxique à l'analyseur sémantique jusqu'à ce qu'il trouve un arbre acceptable par ce module. Le module de dialogue passe alors la liste des propositions au résolveur de références.

Si les références peuvent se résoudre, et s'il n'y a qu'une solution, la liste d'actions est passée au planificateur et au gestionnaire d'actions. S'il y a plusieurs possibilités, la situation est clarifiée par le gestionnaire de dialogue et ceci est signalé à l'utilisateur par un message oral.

Nous avons implanté une stratégie simple pour gérer les clarifications multimodales en utilisant le coefficient de focus. Le gestionnaire de dialogue demande une désignation par pointeur et le processus de référencage est répété avec la dernière interprétation sémantique. L'objet pointé est alors désigné sans ambiguïté puisqu'il possède la plus haut coefficient de la liste. S'il n'y a pas de solution au référencage, et s'il reste des arbres d'analyse syntaxique, le gestionnaire de dialogue prend une autre interprétation sémantique et la passe au module de référencage jusqu'à ce que la liste d'arbres d'analyse soit vide. Si rien n'est trouvé, on signale l'échec à utilisateur.

Si l'action est exécutable – correspond à une commande de navigation implantée dans le système – le gestionnaire d'actions enchaînera et déplacera l'incarnation de l'utilisateur. Sinon, s'il y a échec, le gestionnaire de dialogue recommencera le

processus avec l'arbre d'analyse suivant. L'exemple d'un tel échec est donné par *prends de la hauteur*. Le système interprète cette phrase d'abord comme un verbe transitif qui ne peut pas être exécuté car l'agent n'est pas doté de capacités de manipulations, puis comme une locution (*s'élever*).

Lorsqu'un énoncé correspond à une commande exécutable, le système l'acquiesce oralement en émettant un message pendant qu'il l'exécute. Ce message est tiré au hasard parmi un ensemble réponses positives : *d'accord, oui*, etc. Sinon, une fois que toutes les possibilités d'analyse sont épuisées, le gestionnaire de dialogue rejette l'énoncé en en indiquant la cause. Le générateur de phrases utilise des messages type et en choisit un au hasard. Le générateur peut créer des messages incorrects du point de vue syntaxique. Pour les éliminer, il applique l'analyseur syntaxique sur ses réponses. Si cette analyse syntaxique échoue, le générateur tirera au hasard un autre message type.

4.5.7 La planification et la gestion des actions

Le module de planification et de gestion d'actions accomplit le mouvement de l'agent dans le monde virtuel. La planification est relativement générique et permet d'animer différents modèles de comportement. Pour ceci, on définit les incarnations de l'agent dans le monde virtuel, par exemple un humain ou une voiture, par un modèle géométrique hiérarchique avec ses articulations. Un humain aura un corps qui sera constitué d'une tête, d'un torse, de deux membres supérieurs, de deux membres inférieurs. Un membre inférieur ayant une cuisse, une jambe et un pied (Figure 24).

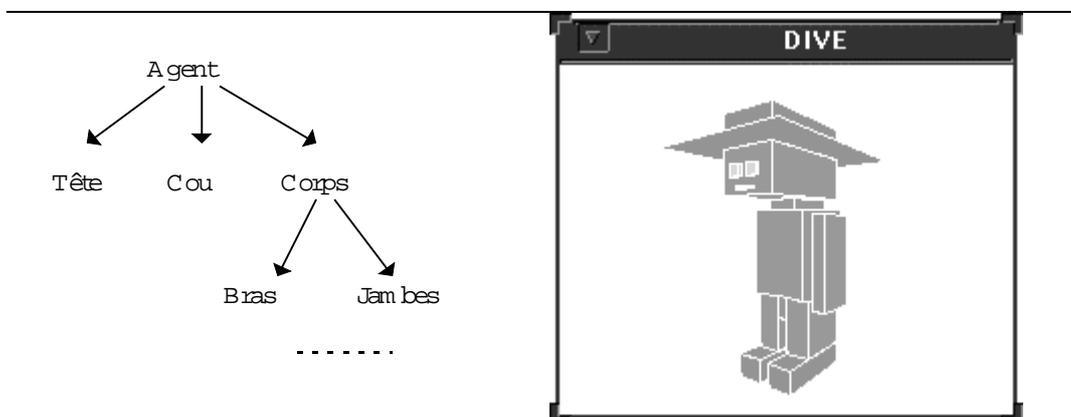


Figure 24 La décomposition d'un avatar humain.

La planification exploite les résultats produits par le raisonneur géométrique. Elle convertit la liste d'actions tridimensionnelles du module sémantique en une séquence de coordonnées de points qui tracent le déplacement de l'agent. Pour ceci, on utilise un algorithme de planification linéaire hiérarchique.

Chaque forme logique du module sémantique est reliée à un plan. Elle est décomposée récursivement en une liste de sous-plans parallèles ou séquentiels. Le comportement et les mouvements possibles d'un agent sont modélisés sous la forme

d'actions décrites comme des règles. Le but du planifieur est de réduire la règle initiale en une suite de règles plus simples, jusqu'au moment où on peut réaliser l'animation physique.

Les règles sont semblables au format classique de STRIPS (Fikes & Nilsson 1971) : elles ont une partie précondition qui doit être vérifiée pour que la règle se déclenche. Cette dernière peut correspondre à un plan exécutable (terminal) qui sera ajouté à la liste des actions à réaliser dans le monde ou bien elle se décompose dans d'autres problèmes – de nouvelles règles – qui seront ajoutées et devront être résolues. Chaque règle contient enfin des faits à ajouter qui traduiront les modifications consécutives à leur action.

Le planifieur a trois plans initiaux qui correspondent aux trois verbes de base : *aller*, *regarder* et *se tourner*. Les trois verbes se réalisant en parallèle pour par exemple aller dans une direction tout en tournant la tête vers une autre. Le plan initial *aller* sera d'abord décomposé en une suite d'actions permettant d'éviter les collisions avec les objets statiques. Puis, le planifieur utilisera les règles de comportement et entamera sa décomposition jusqu'à trouver des animations élémentaires exécutables. Un humain qui marche avancera une jambe, puis l'autre en battant des bras jusqu'à ce qu'il atteigne sa destination.

Pour que les actions d'un plan puissent s'exécuter rapidement et en parallèle, le module compilera le plan en une séquence d'unités indépendantes. Chacune de ces unités est constituée d'opérateurs (translation ou rotation) qui décriront des animations simultanées. Comme résultat, on obtiendra les paramètres de mouvement des objets composant la hiérarchie de la représentation.

Enfin, un simulateur cinématique réalisera l'animation pour que le mouvement résultant soit fluide. Il utilisera une fonction de rappel déclenchée tout les 1/4 de seconde qui décomposera la suite d'actions précédentes en une suite de positions intermédiaires. L'incrément de décomposition dépendra de la vitesse de rappel et cette vitesse déterminera le réalisme final (Figure 25).

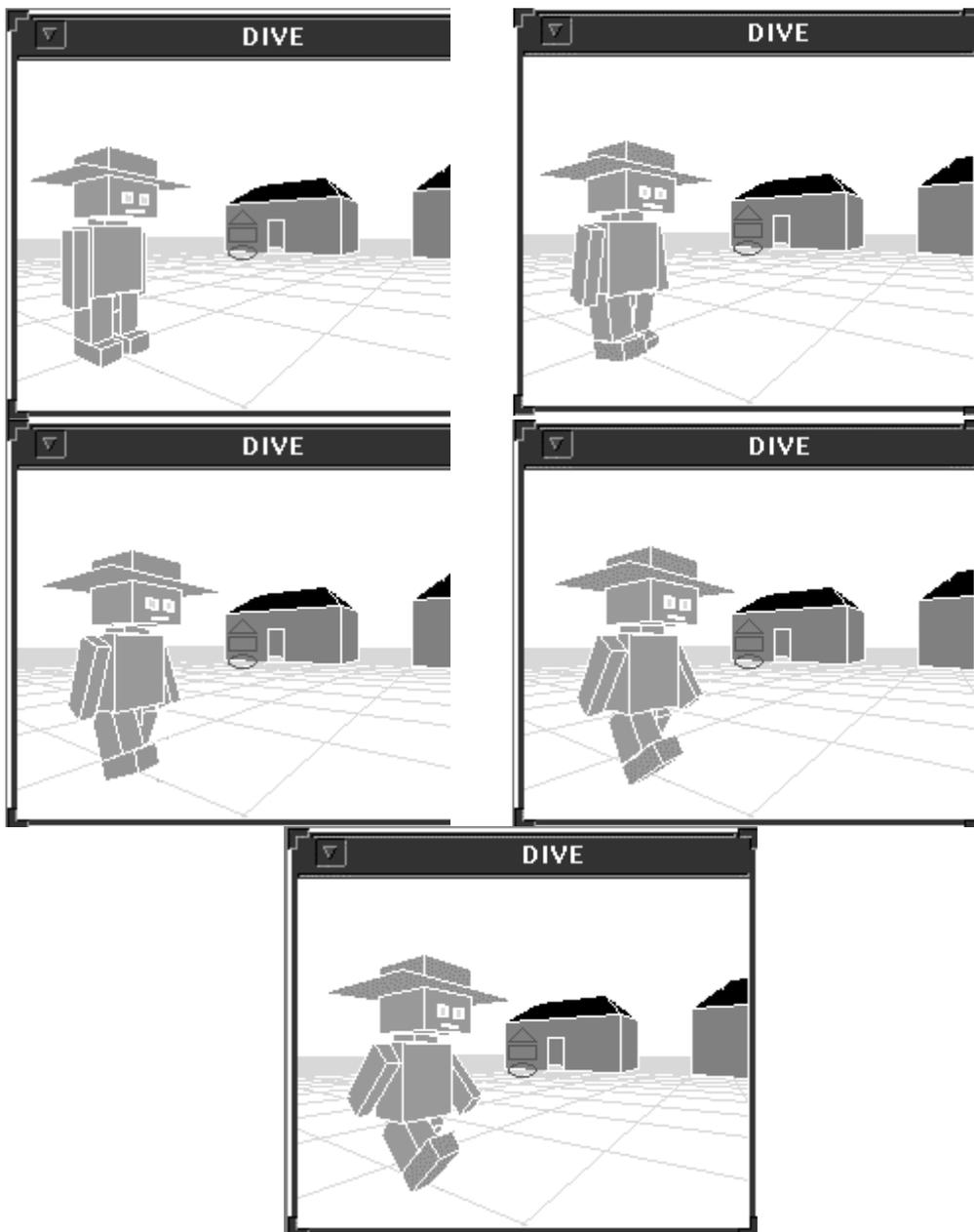
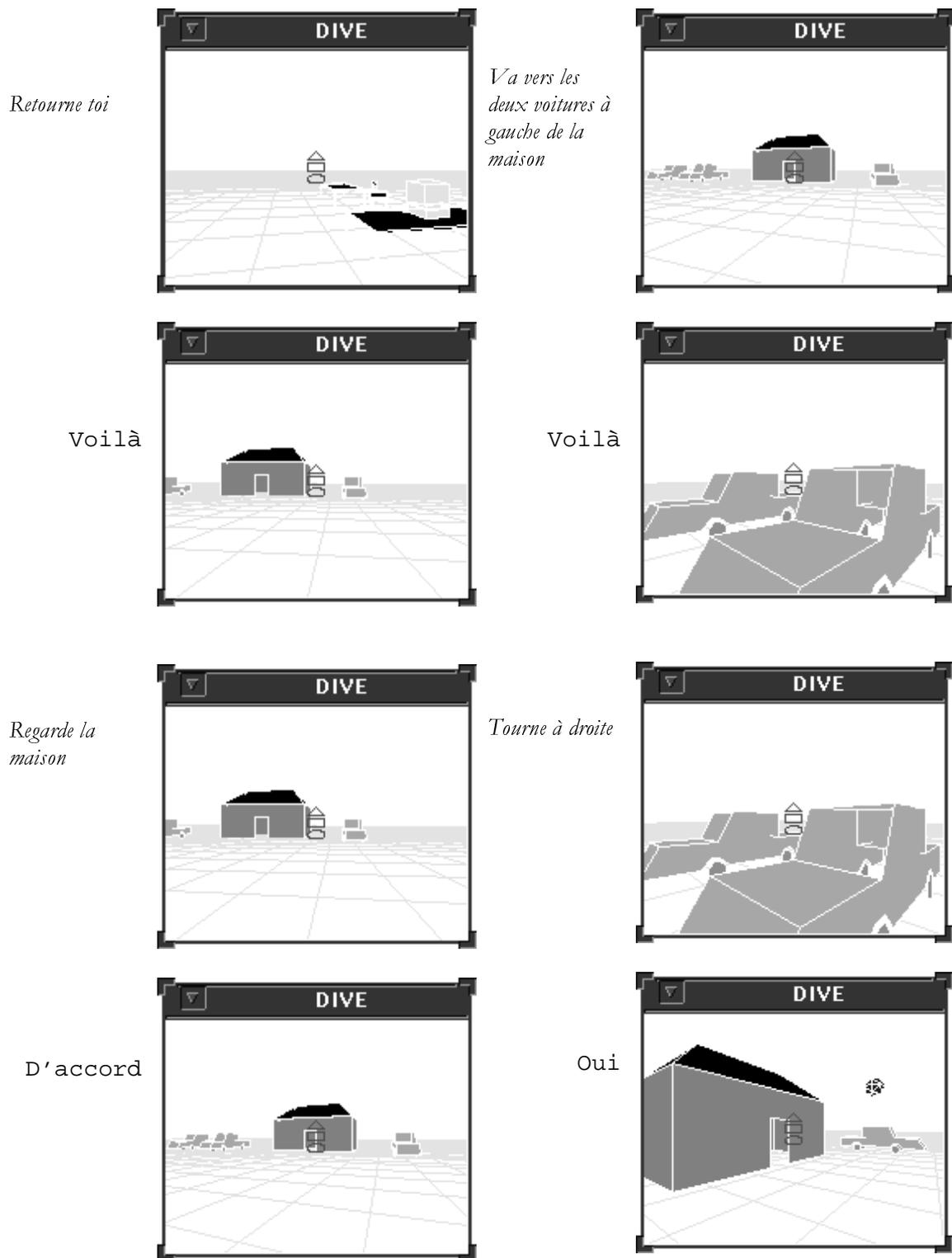


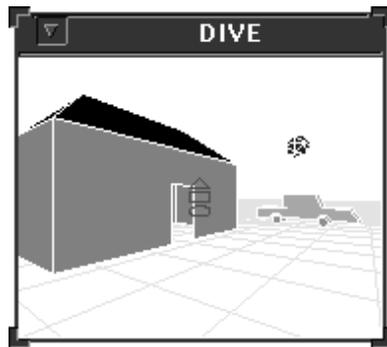
Figure 25 Une animation.

4.5.8 Un voyage avec Ulysse

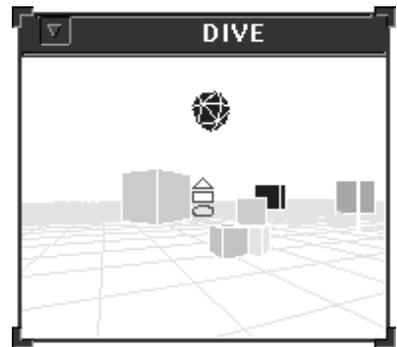
La Figure 26 donne un exemple de dialogue possible avec Ulysse.



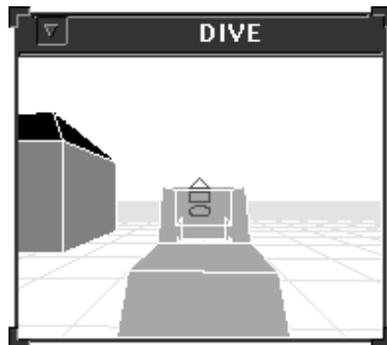
*Va devant cette
voiture*



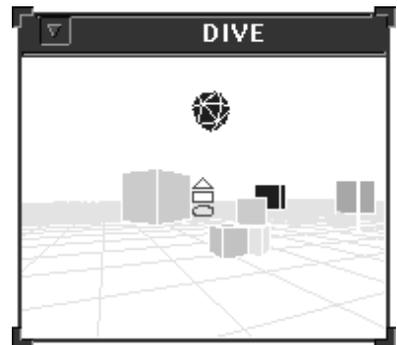
Va vers le cube



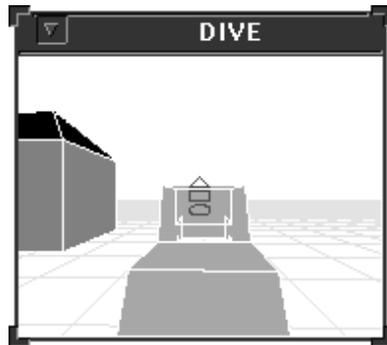
Voilà



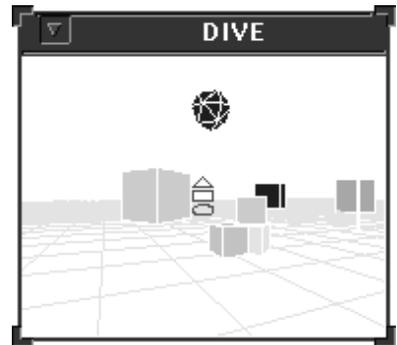
Il y en a
plusieurs



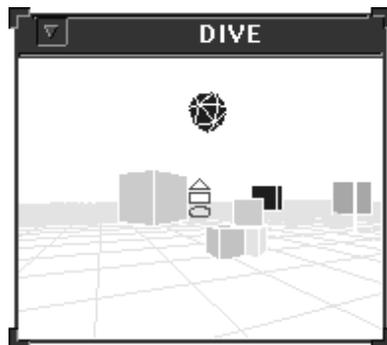
Tourne à droite



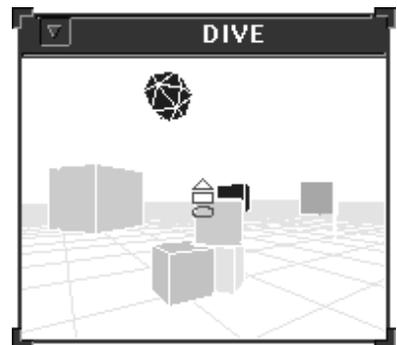
*Va vers les
petits cubes*

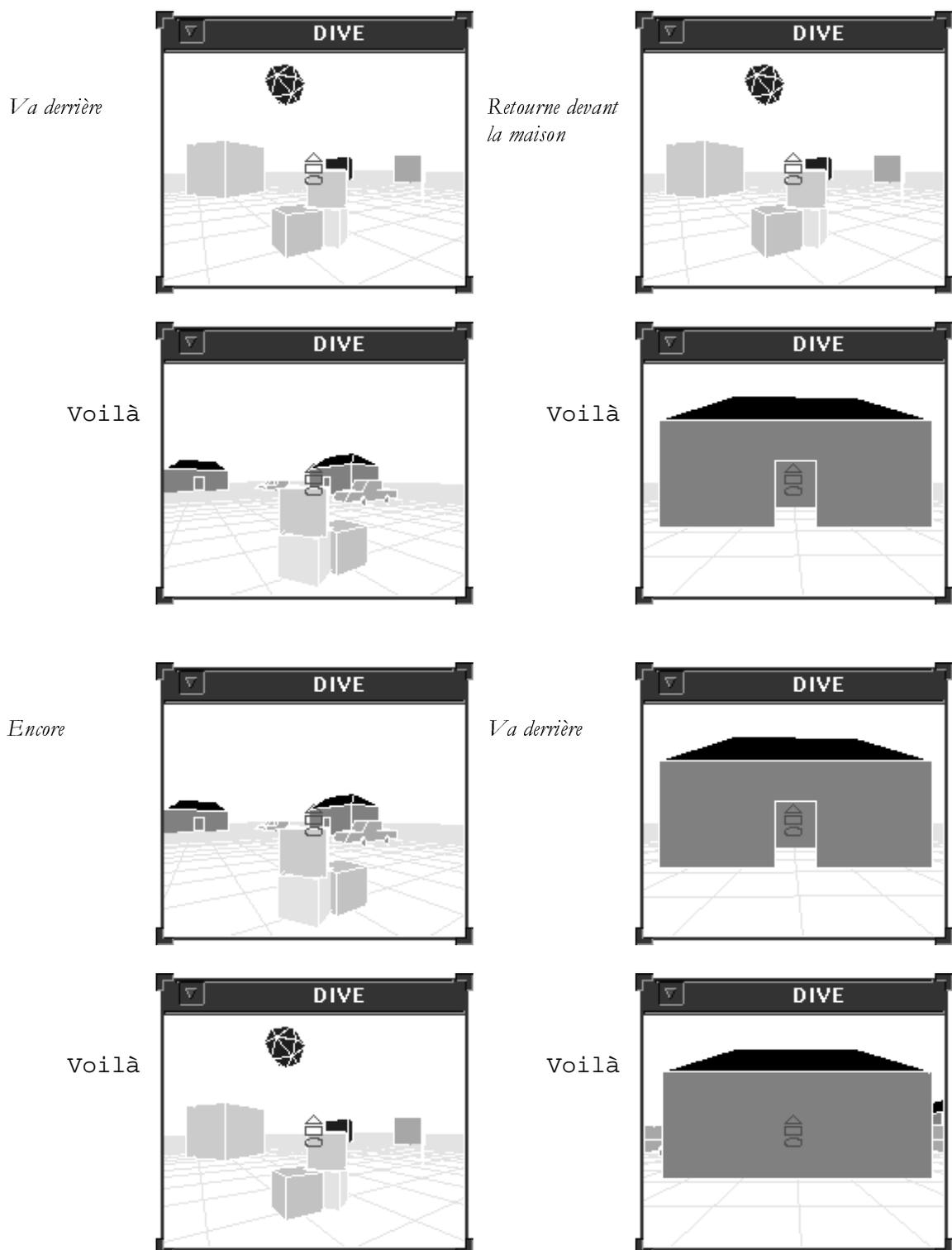


Voilà



Voilà





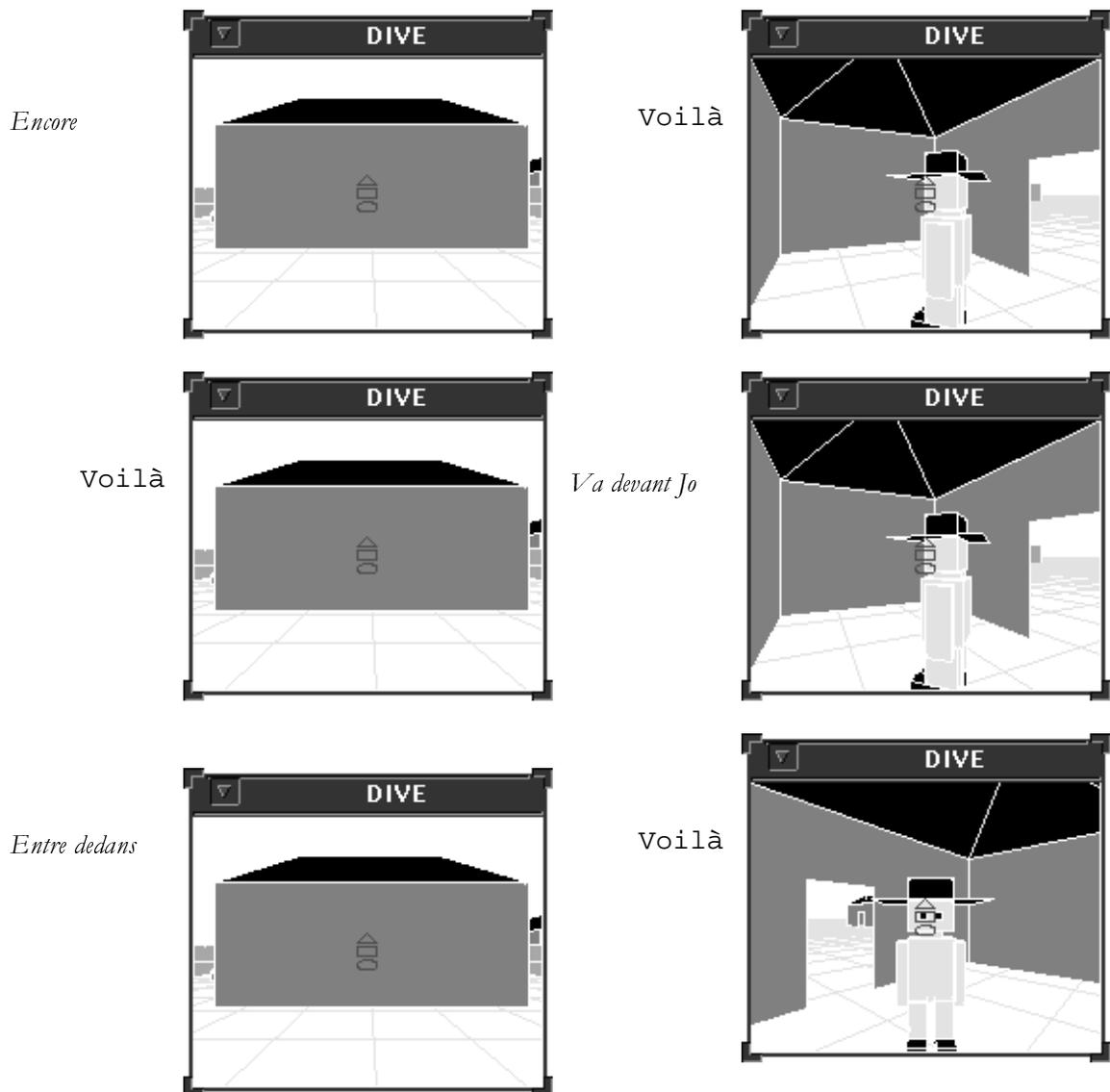


Figure 26 Un voyage avec Ulysse.

4.6 L'adaptation d'Ulysse à la navigation dans le cerveau

L'architecture d'Ulysse a été conçue au départ pour naviguer dans un monde de bureaux. Nous l'avons adaptée à la navigation orale dans des images tridimensionnelles de cerveau. Ces images ont été obtenues par résonance magnétique nucléaire (RMN) et elles ont été fournies le Centre François Baclesse de Caen. Le travail d'adaptation a été réalisé dans le cadre du stage de fin d'études d'un étudiant allemand – *Diplomarbeit* – (Ludwig 1996, Ludwig et al. 1997).

Pour réaliser ce projet, nous avons travaillé avec le groupe artistique *Das synthetische Mischgewebe*. L'idée était de combiner des images réalistes avec des dialogues appropriés pour explorer certaines régions du système nerveux et leurs fonctions. *Das synthetische Mischgewebe* voulait réaliser une performance mêlant l'art et la technologie.

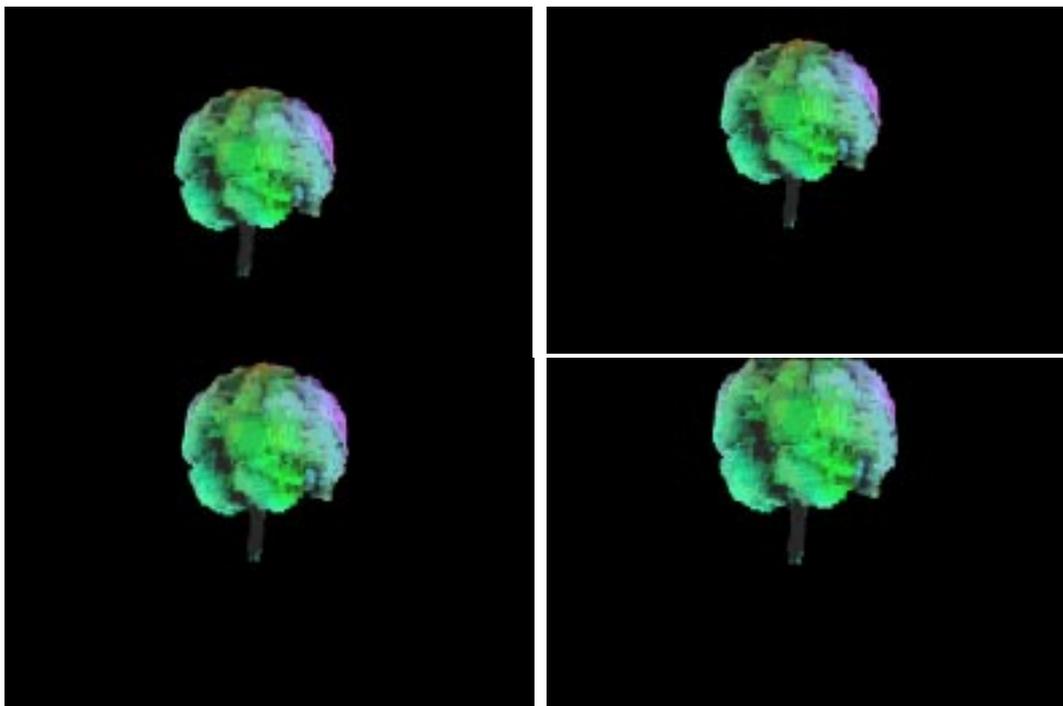
Avec eux, nous avons conçu des scénarios de navigation et d'interaction dans ce nouveau monde virtuel qu'est un cerveau reconstruit. Nous avons limité les scénarios d'exploration aux régions principales du cerveau, c'est-à-dire, les hémisphères, les lobes frontaux, temporaux, le tronc, ainsi que certaines sous régions, telles que l'amygdale.

Des objets tridimensionnels volumiques, tels qu'un cerveau, ne peuvent pas se manipuler en temps réel dès qu'ils ont un certain poids. Nous avons réduit l'image tridimensionnelle à des surfaces (Lenoir et al. 1996). Ceci permet une animation sur un PC de plusieurs images par seconde et donc suffisamment rapide pour une manipulation en temps réel. L'interactivité étant un point-clé du réalisme.

L'agent conversationnel et le programme de visualisation ont été adaptés et implantés sur un PC. À la différence d'Ulysse, qui fonctionnait en tandem sur une station Unix et un PC, l'ensemble du système a tourné sur une même machine.

Nous avons modifié l'architecture pour la rendre plus souple. Les différentes entités d'Ulysse ont été divisées en modules indépendants et pilotées comme des fils de commandes de l'agent – des threads –. Nous y avons intégré le moteur de reconnaissance vocale VoiceType (IBM Speech Engine 1996). L'agent accède aux mots reconnus directement par les interfaces de programmation API (IBM 1996).

Le système de navigation dans le cerveau a été montré dans des représentations artistiques. Il a été le centre d'un spectacle présenté par *Das synthetische Mischgewebe* en France et à l'étranger. Les commandes orales permettent des mouvements tels que celui de la Figure 27. Ils correspondent à l'ordre : *je veux voir le tronc*.



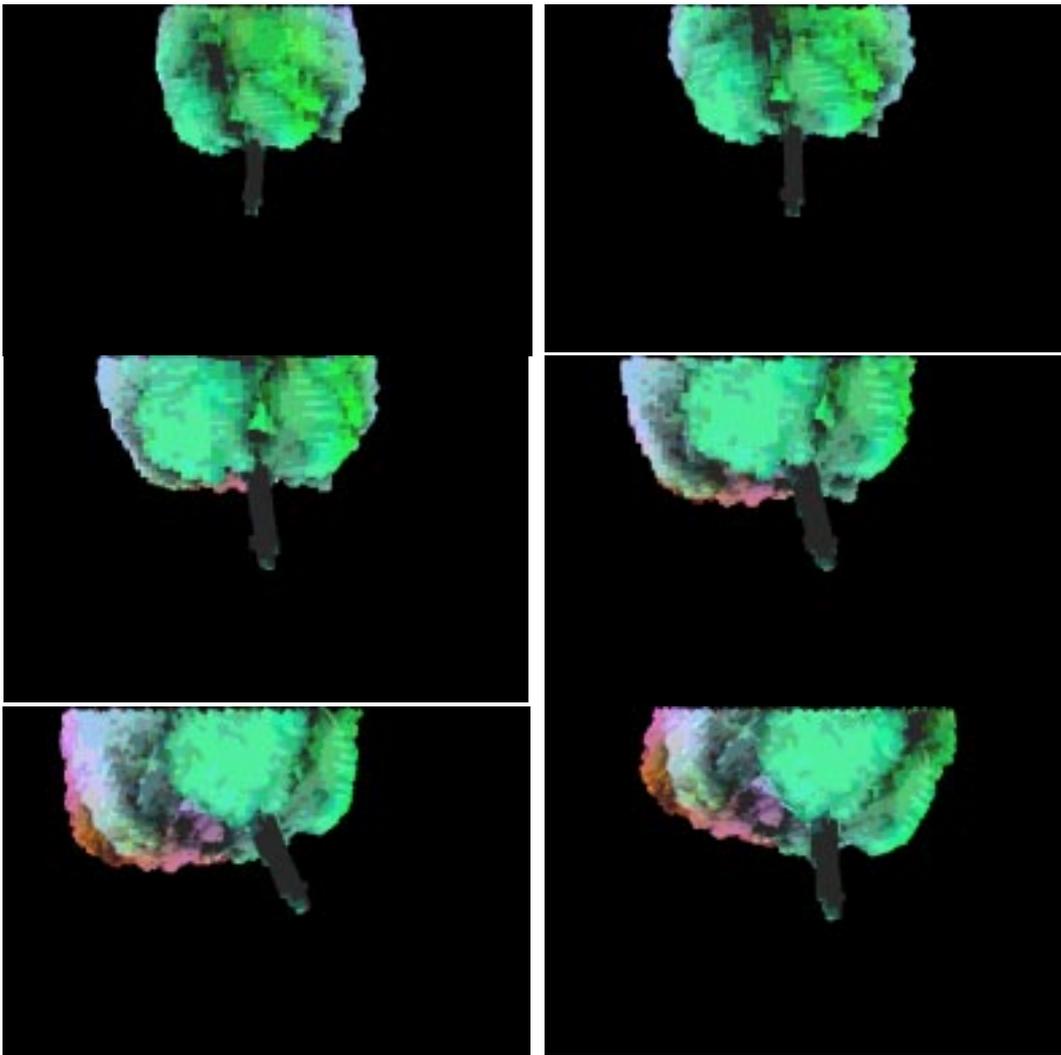


Figure 27 Une animation autour du cerveau.

4.7 Conclusions et perspectives

Dans ce chapitre, nous avons présenté l'agent conversationnel Ulysse. L'architecture d'Ulysse est constituée d'un système de reconnaissance et de synthèse de parole, d'un analyseur syntaxique, sémantique, d'un raisonneur géométrique, d'un résolveur de références, d'un module de dialogue, et enfin d'un planifieur et d'un gestionnaire d'actions.

Ulysse s'incorpore dans la représentation – l'incarnation ou l'avatar – d'un utilisateur du monde virtuel. L'utilisateur s'adresse oralement à cet agent pour réaliser en son nom des mouvements ou des déplacements assez complexes. Nous avons effectué des démonstrations d'Ulysse dans un certain nombre d'occasions,

notamment pour la Science en fête à deux reprises en 1995 et 1996. Nous avons pu constater sa bonne robustesse auprès des enfants.

Ulysse montre qu'une interface linguistique peut améliorer l'utilisation et l'interaction avec des mondes virtuels. Elle facilite la navigation : des déplacements difficiles à réaliser dans un monde virtuel avec une souris, se réalisent facilement à la voix.

Au départ, nous avons mené l'étude de la navigation dans le monde Ithaque (Figure 21). Ce monde représentait un environnement de bureaux. Il était composé de maisons, de voitures, d'un petit mobilier, etc. Nous avons ensuite adapté Ulysse pour naviguer dans des images de cerveau. Pour ceci, nous avons utilisé des images de résonance magnétique nucléaire traitées pour en extraire les surfaces. Le prototype a été montré au cours de représentations artistiques.

Lors de l'adaptation, nous avons un peu modifié l'architecture. Nous avons fait évoluer l'organisation rectiligne du premier prototype vers un système plus modulaire. Elle a permis d'intégrer la reconnaissance vocale au cœur de l'agent par l'intermédiaire des fonctions du module de reconnaissance. Dans ce sens, l'architecture d'Ulysse se rapproche maintenant de celle DictaMed et plus loin dans le passé de celle du système d'interprétation d'électrophorèses. Cette architecture est relativement générique et son adaptation à une nouvelle machine et à un nouveau domaine en a montrée la souplesse.

Ulysse ouvre des perspectives de recherche pour le dialogue, l'informatique linguistique et la description des objets. Lorsque nous testons le prototype, il arrive souvent que nous découvrons des nouveaux besoins pour de nouvelles actions de navigation ou de description d'objets. Ces manques ouvrent notamment la voie à une étude plus approfondie du lien entre la langue et l'espace et du raisonnement spatial. De ce point de vue, les systèmes de réalité virtuelle offrent un cadre d'expérimentation unique et permettent de sortir d'une théorisation fondée sur l'introspection.

Au-delà de la navigation, Ulysse pourrait être doté de facultés pour manipuler les objets, par exemple, ou pour produire un raisonnement plus complexe. L'ajout de telles facultés sort alors du domaine proprement linguistique pour rentrer dans celui du raisonnement et du lien entre ce raisonnement et la langue. Avec l'apparition de mondes virtuels sur le réseau, de telles interfaces aideront à préciser ce lien. Elles aideront aussi à découvrir, explorer et maîtriser ces nouveaux espaces de travail, de culture, de jeux et de formation.

5.
Conclusion

J'ai présenté, dans ce dossier d'habilitation à diriger des recherches, les principaux domaines dans lesquels j'ai eu l'occasion de travailler : les systèmes de raisonnement pour l'interprétation d'images et l'informatique linguistique. Cette contribution m'a permis de dégager une nouvelle problématique et de tracer un programme de recherches à venir. C'est l'objet de ce chapitre.

5.1 Contributions et problématique

5.1.1 Systèmes de raisonnement pour l'interprétation d'images

Dans le domaine de l'interprétation d'images, j'ai proposé et implanté un système d'analyse de gels d'électrophorèses bidimensionnelles. Ce système a mis en œuvre une analyse locale et partielle d'un gel. On peut ainsi identifier des protéines connues sur une image isolée. Cette technique s'oppose à d'autres qui comparent et référencent les protéines à l'intérieur d'une série de gels. Le système d'interprétation suppose alors qu'on dispose d'une connaissance des protéines à rechercher, ainsi que d'une méthode de raisonnement pour les reconnaître.

L'électrophorèse bidimensionnelle, qui est d'une manipulation difficile, suscite un intérêt plus grand maintenant. Elle donne naissance à de nombreuses bases de données accessibles sur le Web. Ceci est la conséquence de l'étude du génome qui produit une quantité considérable de gènes séquencés. La plupart de ces gènes ont une fonction encore inconnue. L'effort de déchiffrement se porte de plus en plus sur l'étude du « protéome », c'est-à-dire l'observation du comportement des protéines d'un tissu (Nowak 1995, Kahn 1995).

On observe le comportement des protéines en modifiant quelques éléments du code génétique ou en faisant subir à des tissus des variations chimiques, thermiques, etc. L'électrophorèse bidimensionnelle permet de tracer la signature protéique d'un tissu. On constate immédiatement les protéines apparues, disparues ou modifiées. On peut ainsi contrôler la toxicité de médicaments, de produits chimiques ou de conditions d'environnement.

Le système que j'avais conçu lors de ma thèse permet d'attacher à un gel la connaissance de l'emplacement de certaines de ses protéines et de les retrouver automatiquement. Il permet aussi d'incorporer d'autres connaissances sur chacune des protéines. On peut imaginer que les efforts de déchiffrement permettent d'aboutir à la liste exhaustive des protéines des tissus et ainsi pouvoir étiqueter entièrement un gel. L'orientation que j'avais prise, qui apparaissait incertaine en 1988, semble maintenant confirmée comme une direction vers cette cartographie.

Le système d'interprétation d'électrophorèses a reposé sur une architecture permettant de combiner des sources de connaissances de traitement d'images et de raisonnement pour traiter l'image du gel. Il a eu comme suite un système similaire pour recalculer automatiquement des images de radar sur des cartes géographiques.

Cependant, à la différence des électrophorèses, le raisonnement géométrique a utilisé une gestion d'hypothèses pour procéder au recalage.

5.1.2 L'analyse syntaxique

J'ai abordé l'analyse syntaxique lors de la conception d'un système de dialogue oral, puis lors de son extension à la langue écrite. Dans ce cadre, j'ai eu la direction d'étudiants en DEA et en thèse. Nous avons démarré ce projet en utilisant un modèle de la phrase entière constitué de règles syntagmatiques. Ce type de modèle entraîne la recherche d'une structure globale. En parallèle, nous avons construit un atelier de développement de l'analyse syntaxique utilisant une structure de chart avec des mécanismes d'unification.

Nous nous sommes attachés à traiter un corpus de comptes rendus médicaux dont les phrases étaient de taille et de style très variés ainsi qu'un autre composé d'ordres de déplacement et de mouvements. L'analyse globale a été possible pour le corpus d'ordres dont les phrases étaient de longueur courte. En revanche, elle a échoué pour les phrases plus longues du corpus de comptes rendus médicaux.

Nous avons évolué vers une démarche en couches pour traiter d'abord les niveaux jugés plus simples de la syntaxe – les groupes – et pour ensuite remonter vers les niveaux plus complexes de dépendance entre les groupes. L'ensemble détecte premièrement les locutions puis il exploite des modèles syntagmatiques pour détecter les groupes nominaux et verbaux, des règles pour déterminer les bigrammes valides de groupes et des règles enfin pour établir les dépendances entre ces groupes.

Le système utilise une architecture en couches et un raisonnement hypothétique à base de chart. Bien que le domaine d'application soit différent, l'architecture et les méthodes ont de nombreux points de contact avec les systèmes d'interprétation d'images décrits auparavant.

L'analyseur est muni d'une interface graphique. Elle a permis de développer interactivement des règles et de visualiser les résultats. Nous avons évalué l'analyseur sur plus de cinquante phrases. Grâce à la méthode en couches, nous avons pu parvenir dans la plupart des cas à l'analyse globale complète de la phrase et dans tous les cas à des résultats locaux partiels.

L'histoire et le sens de notre démarche sont semblables à celles d'autres groupes de recherches. Parmi eux, l'exemple le plus connu est certainement celui de FASTUS (Appelt et al. 1993) qui utilise des cascades d'automates. FASTUS a été créé pour analyser des dépêches d'agence de presse dans le cadre des concours MUC. Depuis, il sert d'exemple à de nombreux groupes de recherches et il a modifié de façon notable les méthodes d'analyse syntaxique dans le monde anglo-saxon.

À la différence de FASTUS, notre système peut gérer des hypothèses. Ceci permet de construire un arbre complet. En ce sens, notre système d'analyse est voisin de celui de Collins (1996) qui fait appel à des statistiques pour produire l'arbre final d'analyse. Cette capacité de gestion d'hypothèses est importante lorsqu'on utilise une stratégie ascendante locale, notamment dans le cas de l'oral. Elle rend le système plus souple en lui évitant de s'engager complètement dans une voie de manière prématurée. Elle lui permet de réviser et d'améliorer son résultat en cours d'analyse.

5.1.3 L'interaction linguistique dans un monde virtuel

L'interaction linguistique dans un monde virtuel est le dernier domaine de recherche auquel je me suis consacré. Il a eu pour cadre un projet européen COST sur le travail coopératif à l'aide de l'ordinateur. Comme pour l'analyse syntaxique, j'ai dirigé des étudiants en thèse et en DEA pour ce projet. Nous avons conçu l'agent conversationnel Ulysse pour résoudre l'un des points les plus difficiles de l'interaction avec les mondes virtuels : la navigation de l'utilisateur dans un monde.

Simultanément à la construction du prototype de navigation Ulysse, nous avons collecté des corpus de dialogue pour préciser le vocabulaire des utilisateurs et les modes de communication. Ulysse se compose de différents modules chargés de détecter les locutions, d'analyser la syntaxe, de construire une forme logique, de résoudre les références et raisonner sur la géométrie de la scène.

La brièveté des énoncés a permis l'utilisation d'une grammaire syntagmatique. La construction de la formule logique dérive d'une grammaire de cas. La résolution des références utilise une méthode de saillance. Le raisonnement géométrique intègre les types des verbes et des prépositions. Il tient compte de la forme et de l'emplacement des objets. Enfin, nous avons construit un planificateur qui réalise le déplacement de l'incarnation d'un utilisateur en évitant les obstacles.

Nous avons d'abord organisé les modules dans une architecture rectiligne dans le cadre d'un premier prototype permettant la navigation dans un monde de bureau. Nous avons ensuite séparé les modules linguistiques au sein d'une architecture plus modulaire lors de l'adaptation à la navigation dans un cerveau reconstruit. L'architecture s'est alors rapprochée de celle des systèmes d'interprétation d'images que j'avais conçus auparavant.

Les interfaces orales pour des environnements virtuels sont encore assez rares. Elles suscitent parfois le scepticisme aussi bien de la part de la communauté des linguistes que de celle de la réalité virtuelle. Nous avons mis en évidence leur complémentarité. L'image est essentielle pour comprendre et elle complète la langue. De plus, il est souvent plus facile de montrer que de décrire :

Was gezeigt werden kann, kann nicht gesagt werden⁵,

À l'inverse, les interfaces linguistiques peuvent faciliter de manière très notable l'utilisation d'environnements virtuels. Une interface orale améliore la sensation d'immersion, ainsi que les capacités de navigation et d'interaction.

Ulysse est un agent conversationnel qui s'intègre dans un monde virtuel. Avec lui, l'utilisateur dispose d'une interface linguistique et visuelle tridimensionnelle. Leur combinaison permet de s'affranchir de certaines limites de chacune des interfaces et de tirer parti de leurs avantages respectifs.

⁵ Ludwig Wittgenstein, 1922, § 4.1212, (*Ce qui peut se montrer ne peut pas se dire*).

5.2 Recherches futures

5.2.1 Les enjeux et les objectifs

5.2.1.1 L'Internet

L'Internet et le *World Wide Web* ont rendu disponible une masse considérable de données accessibles presque instantanément. En informatique et dans d'autres disciplines, les universités du monde entier ont mis, dans un même mouvement et en quelques mois, les descriptions de leurs projets, leurs rapports de recherches, leurs programmes informatiques et même parfois leurs données les plus essentielles, à la disposition de tous. Ces informations sont en grande partie sous la forme de textes. Elles se composent aussi de plus en plus d'images.

L'Internet permet, d'autre part, des échanges plus sélectifs d'informations entre groupes et entre personnes, soit par l'intermédiaire de réunions, soit par l'envoi direct. Il est ainsi un outil de coopération entre des utilisateurs connectés. La coopération se réalise de façon synchrone par des groupes de discussion en temps réel (*chat*) et par la visioconférence ou bien de façon asynchrone par le courrier électronique et par des services de nouvelles (*news*).

L'Internet permet de partager et d'échanger de la connaissance en réduisant le coût de transaction à presque rien. Il est devenu banal d'affirmer que l'extension de ce réseau aura pour conséquence vraisemblable une transformation de l'organisation des entreprises, de l'éducation et peut être de la société.

L'Internet n'aurait que des avantages si le flot d'informations qu'il créait ne devenait pas très vite inexploitable à cause de sa quantité. Parfois aussi, l'inadaptation de ses interfaces rend son accès difficile.

C'est dans le cadre général d'élaboration de nouveaux concepts et de nouvelles techniques, à la fois pour structurer l'information, les connaissances et les schémas de collaboration, que je fixe mon domaine de recherches futures. Mon sentiment est que de telles techniques sont notamment du ressort de la visualisation et de l'exploration. Conjointement, elles peuvent aider l'utilisateur à construire une représentation mentale de l'Internet et de ses outils et, ainsi, en améliorer leur compréhension.

5.2.1.2 Les domaines de travail

Les systèmes de réalité virtuelle portent en eux la représentation visuelle et l'interaction. Ils réalisent une projection immédiate de la connaissance dans l'espace et offrent la possibilité de l'explorer et de la manipuler de façon dynamique. Cependant naviguer dans des espaces virtuels et interagir avec les objets de tels espaces restent difficile. Peu de chercheurs ont exploré jusqu'à présent l'interaction linguistique dans les environnements virtuels. J'ai montré par un travail précédent qu'elle peut contribuer à résoudre les difficultés de navigation.

L'objectif de recherche que je propose maintenant prend appui sur le travail passé et notamment sur l'agent Ulysse qui en a cristallisé une partie sous la forme d'un prototype. Ce travail futur aura pour but de développer de façon plus systématique les interfaces linguistiques et les agents conversationnels dans les mondes virtuels.

Des nouvelles formes d'agents matérialiseront ces recherches. Elles devront pouvoir s'affranchir des normes de description des mondes et avoir des capacités linguistiques et de compréhension couvrant des variétés de contextes beaucoup plus nombreuses. Ce travail devrait permettre aussi de mieux établir les mécanismes qui relient la langue, l'espace et l'information.

Le plan de recherches qui correspond à ces projets peut se diviser en trois domaines d'action distincts :

- élaborer des nouveaux mécanismes pour permettre :
 1. l'exploration d'un grand nombre de mondes virtuels,
 2. l'interaction avec les objets des mondes et
 3. la délégation de tâches dans ces mondes
- développer des processus d'acquisition automatique de connaissances linguistiques, que celles-ci aient trait à la syntaxe, à la sémantique, au raisonnement ou au dialogue
- relier ce travail à des études cognitives pour mieux comprendre le rôle de la langue dans la description spatiale et la construction d'espaces d'informations mentales.

Le travail fixé devrait permettre de dégager de nouveaux paradigmes d'interactions tridimensionnelles et de préciser la fonction que la langue peut y jouer. Il devra aussi aboutir simultanément à un ensemble de réalisations concrètes qui associent des interfaces linguistiques à des mondes virtuels ainsi qu'à de nouveaux prototypes d'agents conversationnels. Cette double démarche est essentielle dans mon esprit. En effet, les réalisations permettent de se confronter à des phénomènes et des difficultés qu'il est difficile d'imaginer *a priori*. Elles mettent en évidence une autre problématique qui aboutit en retour à créer d'autres concepts et d'autres réalisations. Elles devraient aussi aider à trouver des débouchés concrets sous la forme de nouveaux systèmes d'enseignement et de transferts vers l'industrie.

5.2.2 Vers des agents conversationnels aux capacités plus étendues

Notre première perspective consiste à déployer de façon systématique des agents conversationnels dans des mondes virtuels. Ceci passe par l'étude de nouveaux mécanismes pour développer les modules de l'agent et pour le doter de facultés plus étendues.

Nous nous proposons d'utiliser, au départ, les concepts dégagés par Ulysse et de les étendre à d'autres mondes et d'autres tâches. Les facultés des agents devront permettre à l'utilisateur d'explorer et d'interagir avec un grand nombre de mondes virtuels, de commander des dispositifs ou d'autres agents dans ces mondes et enfin de déléguer à l'agent des tâches à réaliser dans les mondes.

5.2.2.1 La langue pour explorer des mondes virtuels

Le principe essentiel d'un agent conversationnel dans un monde virtuel, tel que nous l'imaginons, est son incorporation à l'intérieur de la représentation d'un utilisateur ou d'un objet (un artefact). On attribue ainsi à cette représentation des facultés, des pouvoirs supplémentaires, dans le monde. Ce modèle est déjà présent dans Ulysse. Bien que limités pour l'instant à quelques mondes, ses modèles

d'interaction linguistique et spatiale, ainsi que sa structure de dialogue et son plan de communication, lui permettent de voyager.

Un ensemble grandissant de mondes virtuels, facilement accessibles, rendent possible l'extension du champ de navigation d'Ulysse. Cependant, la variété des mondes virtuels va de pair avec l'augmentation du nombre des objets potentiels et la complexité de leur description. Cette variété impose un enrichissement du vocabulaire, de la syntaxe et de la sémantique pour que l'agent puisse prendre en compte les nouveaux contextes. Elle entraîne aussi de sa part des actions plus élaborées.

La diversité potentielle des actions de navigation imposera une refonte des algorithmes de planification d'actions ainsi que d'animation et d'articulation des mouvements. L'agent devra faire en sorte que le déplacement de l'utilisateur, ou celui d'autres avatars, dans un monde virtuel soit ressenti comme plus « naturel ». Il devra prendre en compte la modification constante du monde et d'éventuels événements qui surviendraient pendant son action. Ceci facilitera l'interaction entre d'une part l'utilisateur et sa représentation et d'autre part l'utilisateur et des entités d'un monde virtuel. Ces entités pouvant représenter des êtres humains ou des agents.

Les mondes virtuels ont des domaines de représentation très vastes. Pour ce qui concerne la construction d'un agent d'exploration, nous nous limiterons dans un premier temps à des mondes architecturaux et anatomiques.

S'affranchir des normes de description des mondes virtuels

Lors de la construction d'Ulysse, nous avons utilisé, comme monde d'expérimentation, une scène représentant des maisons, des bureaux et des voitures. Nous avons ensuite adapté Ulysse à la navigation dans le cerveau. À l'époque du démarrage d'Ulysse, il y avait peu de mondes facilement disponibles.

La situation a complètement changé et il existe désormais un assez grand nombre de mondes virtuels. Certains annuaires en recensent des centaines dans des domaines très variés. Ces mondes utilisent désormais très souvent le langage de description géométrique VRML (*Virtual Reality Modeling Language*). Ce langage apparaît comme une sorte de *lingua franca* de la réalité virtuelle.

Un des annuaires les plus importants est l'archive VRML de l'université de San Diego. Les domaines couverts vont des mondes personnels et de représentation de structures d'entreprise, à l'architecture, l'art, l'astronomie, les sciences biomédicales, la chimie, les applications commerciales, la récréation de mondes historiques, les cartes. Ils comprennent aussi des applications comme la simulation de l'environnement, des exercices interactifs d'informatique, de sciences physiques et de mathématiques. La domination du langage VRML semble irréversible et cette norme est lisible par la plupart des navigateurs (*browsers*) du Web.

Pour permettre à un agent de voyager au long cours dans des mondes virtuels, un premier travail, très concret, est de le rendre compatible à cette norme. Une telle adaptation concerne surtout les parties relevant de la géométrie et des animations. Cette tâche peut paraître en marge d'un travail scientifique de fond. Cependant, il ne faut pas en minimiser l'importance. Elle exige de consacrer à la fois du temps et des efforts mais elle accroît de façon considérable le potentiel de diffusion de résultats dans l'industrie et la communauté scientifique.

Le Plan de Rome et d'autres mondes architecturaux

Nous nous donnons comme objectif suivant d'implanter dans l'agent les capacités pour explorer un grand nombre des mondes architecturaux. Ces mondes pourraient provenir de l'archive de San Diego ou d'autres sites. L'agent devra permettre à un utilisateur de naviguer à la voix et à la souris dans de tels mondes. L'utilisateur devra pouvoir se diriger vers les différents objets, se positionner relativement à eux, les visiter, décrire des mouvements à exécuter, en bref, aller là où il le veut.

Pour chacun des mondes nous créerons le lexique correspondant. Ceci imposera une extraction du vocabulaire que nous obtiendrons en utilisant des sources disponibles telles que des lexiques d'architecture, de génie civil, de mécanique, etc. ainsi qu'en collectant des corpus d'interaction avec des utilisateurs. Comme résultat, nous devrons pouvoir nommer et décrire les différents objets des mondes. Nous compléterons aussi la liste des verbes de mouvement que nous possédons déjà et nous définirons leur sémantique. Nous exploiterons pour ceci des sources lexicales disponibles telles que des dictionnaires de verbes et encore des corpus d'interaction.

D'autre part, l'université de Caen possède une maquette en plâtre de la Rome antique : le Plan de Rome de Paul Bigot. Le Plan de Rome est en cours de numérisation tridimensionnelle sous la direction de Philippe Fleury. C'est une entreprise de longue haleine et très ambitieuse, car elle est doublée d'un souci de précision historique. Elle demande de corriger certaines parties de la maquette en fonction de découvertes archéologiques récentes.

Le Plan de Rome offre un champ d'exploration considérable. Il donne la possibilité recréer une ville qui a fasciné tant d'hommes et de la mettre à la portée de tous. Cependant, sans un guide, il ne sera sans doute pas commode de découvrir la Ville virtuelle – de même que la Ville réelle. Un agent conversationnel de navigation pourrait en faciliter la visite. Il permettrait alors une meilleure compréhension de la maquette virtuelle.

Actuellement, l'équipe de Philippe Fleury a numérisé une demi-douzaine de monuments, isolés les uns des autres. En même temps que l'étude d'autres mondes virtuels architecturaux, nous avons l'intention de doter un agent de capacités lui permettant d'explorer les monuments romains. Ceci pourrait prendre la forme, dans un premier temps, d'un enrichissement du lexique. Il comprendrait les désignations des différentes parties d'un monument et une nouvelle terminologie. Dans un deuxième temps, nous pourrions augmenter les capacités d'actions de l'agent conversationnel au fur et à mesure de la reconstruction de la ville virtuelle.

Des mondes anatomiques

En parallèle avec les mondes architecturaux, nous avons choisi les mondes anatomiques virtuels comme deuxième espace de navigation à la voix. Nous avons déjà adapté Ulysse à une navigation limitée dans un cerveau virtuel. À cette occasion, nous avons aussi donné à l'utilisateur la possibilité d'interagir avec certaines des zones aux fonctions remarquables, telles que la peur, la sueur, etc. Cette navigation a été exploitée dans des représentations artistiques.

Des atlas anatomiques virtuels sont en cours de construction. Ils concernent un certain nombre de parties du corps. Les chercheurs portent un intérêt et un effort tout particulier pour la cartographie précise du cerveau et de ses différentes zones fonctionnelles. Le centre de recherches Cycéron de Caen et des membres de l'équipe

de traitement d'images du GREYC, notamment, participent à ces efforts de cartographie. Par ailleurs, de même que pour l'architecture, des modèles anatomiques de parties du corps sont disponibles en nombre croissant dans le format VRML.

Lors de l'adaptation d'Ulysse, nous avons dû remanier le lexique et introduire de nouvelles dénominations. Elles ont fait intervenir la terminologie médicale appropriée. Le travail mené jusqu'à présent a concerné les dénominations de la surface du cerveau et quelques dénominations internes. Notre objectif est d'enrichir et de structurer le lexique au fur et à mesure que nous disposerons d'objets cérébraux internes et plus précis. Nous pourrions ainsi intégrer un dictionnaire tridimensionnel anatomique. Il s'agira d'associer des mots à des images ou des représentations géométriques pour nous permettre une navigation plus précise et plus souple.

L'extension du vocabulaire pour nommer les zones du cerveau est semblable à celle des mondes architecturaux. Cependant, le cerveau est moins « géométrique » qu'une ville. Les circuits nerveux sont constitués de méandres, de circonvolutions, etc. En relation avec ces objets cérébraux, notre objectif sera d'améliorer la localisation et l'identification de zones à partir d'une description. Ceci pourra faire appel à des prédéfinitions de zones complexes et éventuellement à des procédures de raisonnement utilisant des techniques de reconnaissance des formes.

En corollaire de la dénomination et de la localisation, nous porterons nos efforts sur l'amélioration du mouvement et le suivi des trajectoires. Les trajectoires architecturales sont relativement rectilignes. La navigation dans le cerveau exige parfois le suivi d'une trajectoire plus complexe. Nous isolerons un certain nombre d'entre elles que nous modéliserons et que nous implanterons. Pour ceci, nous espérons pouvoir tirer parti de ce qui a été réalisé dans le cadre de l'interprétation de constats d'accidents de la route et de descriptions de trajectoires automobiles (Enjalbert & Victorri 1994).

Pour cette navigation anatomique, nous avons l'intention de poursuivre notre collaboration avec des groupes artistiques tels que *Das synthetische Mischgewebe* ou des entreprises éducatives. Cette collaboration a entrouvert des possibilités d'applications que nous n'avions pas imaginées. Nous pensons qu'elle continuera d'être fertile. Au delà d'un spectacle, l'intérêt de la navigation à la voix dans le corps humain est aussi pédagogique : elle permet de faire découvrir le corps d'une manière plus attrayante et plus exacte qu'avec des planches. Enfin, il est permis d'espérer qu'elle donne lieu, dans un avenir plus lointain, à des applications cliniques.

5.2.2.2 La langue pour commander

L'incorporation d'un agent dans la représentation de l'utilisateur d'un monde virtuel lui permet de s'y déplacer. De façon complémentaire, nous élaborerons des agents conversationnels pour les intégrer dans les représentations d'autres objets du monde virtuel. L'agent incarné dans un objet disposera d'une interface linguistique. Il sera conçu pour recevoir des commandes des utilisateurs et pour les exécuter. Ces commandes pourront concerner d'une part des manipulations et des déplacements, qui sont assez générales. Elles pourront aussi prendre une forme plus particulière et dépendante des facultés de l'objet.

Les objets d'un monde virtuel peuvent représenter, par exemple, des documents, des agendas, des dispositifs de communication. Un agent incorporé à l'intérieur devra connaître les caractéristiques de ce dans quoi il réside et des tâches que l'utilisateur

peut lui demander de réaliser. Il devra aussi intégrer des capacités de raisonnement, un mode d'exploitation et de gestion et la possibilité d'interagir avec l'utilisateur. De tels agents conversationnels incarnés dans des objets virtuels permettront, nous le pensons, de dialoguer avec eux, de comprendre leur fonctionnement, de les manipuler, de les utiliser plus facilement.

Des manipulations géométriques

Une des applications les plus intéressantes des mondes virtuels réside dans le travail en groupe et plus particulièrement dans la téléconférence. Nous étudierons les agents plus particulièrement dans ce contexte, sans exclusive d'autres applications.

Dans les mondes virtuels coopératifs, les artefacts présents ont souvent pour rôle de symboliser un contexte de travail réel et de produire une impression spatiale. Ils se destinent à rappeler aux utilisateurs un environnement familier et à évoquer des situations connues. On y trouve des salles de réunions avec des portes et dedans des objets tels qu'une table, des fauteuils, etc. C'est le cas des salles de visioconférences de la Figure 17 et de la Figure 18.

Le déplacement et la manipulation spatiale des objets de telles scènes sont symétriques du mouvement de l'utilisateur. Les commandes sont dirigées vers une représentation externe au lieu de se destiner à sa propre représentation. De telles facultés, pour un agent, dérivent d'une structure semblable à celles nécessaires à la navigation.

La manipulation d'objets, qui concerne le placement de structures prédéfinies, pourrait aussi s'étendre à la modification de leur forme et donc au modelage par une description linguistique. Ce problème a des points de contact avec l'animation d'un personnage : l'articulation des membres et des parties qui le composent. Nous avons réalisé ce type d'animations dans Ulysse cependant les contraintes sont là très différentes. Nous examinerons dans quelles conditions on peut étendre la faculté de raisonnement spatial de navigation et d'articulation au modelage de scènes et d'objets.

Dans le cadre du travail en groupe, il est nécessaire d'établir des protocoles de droits et d'exclusivité. Lorsque le monde virtuel est « habité » par plusieurs utilisateurs, certains objets doivent rester personnels. D'autres sont la propriété de groupes et d'autres encore sont accessibles et modifiables par tous les utilisateurs. Il faudra donc en coordonner l'accès ou la modification.

Des modèles de partage ont déjà été étudiés et proposés (Sung & Wohn 1996) et la notion de synchronisation est présente dans certains environnements virtuels partagés. Dans une conférence virtuelle, plusieurs participants peuvent simultanément décider de s'approprier un objet et prendre son contrôle. On doit s'assurer qu'une personne unique est habilitée à le faire. Certains modèles mettent en œuvre des protocoles de synchronisation qui sont voisins de ceux mis en place dans les réseaux et les systèmes répartis. Nous examinerons ces modèles et nous verrons dans quelle mesure, ils peuvent s'appliquer à un ou plusieurs agents.

Doter les objets d'autres capacités

On peut doter les objets des mondes virtuels de facultés plus ou moins complexes. Par exemple les tableaux blancs (Figure 17) répliquent les dessins que tracent un utilisateur sur les différents sites participants. Il est facile d'imaginer que

les mondes virtuels soient peuplés d'autres objets partagés comme des agendas, des documents, etc.

La manipulation virtuelle des objets met en œuvre la sémantique et la géométrie de l'espace. Elle est relativement générique quelque soit le type d'objet. Pour des fonctions particulières à un seul objet, cette sémantique est inopérante. Par exemple, un enregistreur d'images ou de sons, une caméra, dans un monde virtuel devront incorporer une interface qui permette de les désigner et de leur donner des ordres. Ces ordres étant du ressort de leurs capacités, c'est-à-dire de viser une scène, de l'enregistrer pendant une durée définie, de la rejouer.

Dans ce contexte, le niveau syntaxique est le seul qui semble avoir une généralité à l'heure actuelle. Les autres niveaux de sémantique, de raisonnement et d'action doivent sans doute recevoir un traitement adapté à l'objet qui les reçoit. Dans le cas d'une caméra, il s'agit de quelques commandes plutôt simples. D'autres dispositifs voisins ont une sémantique plus complexe, comme par exemple les magnétoscopes programmables, qui font intervenir une notion de temps.

Pour cette partie, qui concerne la conception d'agents de commandes, nous tenterons de préciser les limites de ce qui est particulier à un objet et de ce qui est plus général. Nous choisirons quelques objets courants dans les applications de travail en groupe susceptibles d'être utilisés avec la réalité virtuelle. Ces objets pourront être un agenda, un téléphone et des documents sous la forme de textes. Nous définirons un modèle de communication linguistique. Pour ceci, nous procéderons à des mises en situation d'utilisateurs, en examinant leurs réactions et en collectant des corpus d'interaction.

L'un des enjeux est d'étendre la formalisation sémantique d'un agent conversationnel à la commande d'objets quelconques. Un des points de notre étude sera de les doter d'une représentation d'eux-mêmes pour qu'ils puissent se présenter aux utilisateurs et qu'en retour un utilisateur puisse les questionner ou les actionner.

La modélisation de l'interaction linguistique avec des objets virtuels devrait permettre par la suite d'intégrer des agents conversationnels dans des objets réels. Beaucoup d'objets actuels sont trop complexes pour que les utilisateurs puissent les utiliser avec efficacité. L'incorporation de tels agents aurait pour objectif de simplifier leur maniement.

5.2.2.3 La langue pour déléguer

Rome encore

Les facultés de navigation et de commande conduisent à la possibilité de déléguer; l'agent d'interaction se chargeant de l'intendance. La délégation prendrait la forme dans un premier temps d'une extension de la planification. Il s'agirait de la relier à l'intention de l'utilisateur. La délégation pourrait, pour ceci, être précédée d'un dialogue pour que l'utilisateur précise ses désirs.

Le Plan de la Rome offre de nouveau une application possible pour un tel agent conversationnel. L'agent s'incarnerait comme un guide historique et il accompagnerait l'utilisateur dans la visite de la ville virtuelle. Déléguer consisterait à indiquer à l'agent ce qu'on veut voir, ses goûts, le niveau de détail. Le modèle conversationnel rend possible une négociation au cours de la visite et sa

modification. Des requêtes telles que *Faisons le tour du Forum, puis allons vers la Colisée* seraient possibles.

On peut aussi imaginer d'incorporer des agents dans certains des objets virtuels, voire des personnages du passé, pour leur permettre de se présenter et de se décrire. L'utilisateur pourrait alors interagir avec ces objets et poser des questions telles que *À quoi servait-il? Qui étais-tu? Où es-tu mort?*

À plus long terme, on peut envisager d'associer à un espace virtuel des bases de textes en latin ou dans d'autres langues. On pourrait adjoindre à ces bases des outils de recherche et d'extraction d'informations. L'extraction pourrait tenir compte du contexte spatial de l'utilisateur dans le monde et redonner vie par une citation ou par un commentaire, à un monument, à un Romain.

La gestion partagée

Une deuxième perspective de la délégation est celle de la gestion partagée des agents (*shared control*). La langue, dans un contexte humain, permet de commander un groupe de personnes pour lui distribuer des tâches. Une piste de recherches est d'examiner comment transposer cette capacité, d'abord de façon limitée, puis en cherchant à l'étendre, à un groupe d'agents. La délégation consisterait cette fois dans la division d'une commande globale et dans sa répartition en des sous ordres donnés à chaque agent. Les sous ordres devraient s'accompagner de contraintes de synchronisation.

Une application possible de la gestion partagée est fournie par l'interprétation de constats écrits d'accidents de la route. Ces textes décrivent en général des collisions entre deux voitures. Bernard Victorri et de Patrice Enjalbert dirigent actuellement un projet d'extraction d'informations dans ces constats (Pied et al. 1996).

Une perspective est de produire une simulation de l'accident en faisant appel à un ensemble d'agents conversationnels. Un agent distinct serait alors incorporé dans l'incarnation de chacun des véhicules en jeu. Ensembles, ils rejoueraient la scène. Une partie de la connaissance géométrique pour reconstruire la scène pourrait être fournie par le relevé tridimensionnel de la France par l'Institut géographique national.

Dans ce contexte, une architecture possible serait de faire intervenir un agent de répartition pour analyser le constat et distribuer les ordres aux agents situés dans les voitures. L'agent de répartition accompagnerait ses ordres de contraintes de synchronisation et les voitures devraient s'animer en conséquence.

Certaines actions demanderaient une planification voisine de celle mise en place dans Ulysse. D'autres actions sont rendues nettement plus complexes par la synchronisation. Les voitures devront notamment se coordonner par un « contrôle partagé » de la situation afin d'aboutir à la collision. On pourrait mettre en œuvre un algorithme de planification réactive (Jung et al. 1994, Wilkins et al. 1995), par exemple, qui permette de prendre en charge la poursuite d'un agent par un autre en tenant compte des modifications continues des objets de la scène.

L'analyse des constats d'accidents ouvre la perspective de donner à des agents des comportements autonomes, de relier le contrôle, la gestion, voire la description d'un pouvoir, à une formulation linguistique. Plus concrètement, l'étude de tels agents pourrait permettre d'étudier et de simuler des accidents de la route pour en construire des bases de données exploitables plus facilement et ainsi aider à déterminer ou à isoler les causes qui les provoquent.

5.2.3 L'acquisition automatique de connaissances linguistiques

Les principes linguistiques que nous avons mis en œuvre jusqu'à présent reposent sur un codage manuel des connaissances. Elles s'appuient souvent sur des règles écrites à la main. Ceci est relatif aussi bien à la syntaxe, qu'à la sémantique, au raisonnement et au dialogue. De telles règles nécessitent une introspection de la part de celui qui les écrit. Et il est presque toujours difficile d'obtenir une bonne couverture avec une écriture manuelle et de garder les règles cohérentes.

Notre ambition est de construire un agent pour naviguer et interagir dans une grande variété de mondes. Il n'est pas envisageable de décrire toute la connaissance correspondante à la main. Il sera donc nécessaire de développer des processus d'acquisition automatique de connaissances linguistiques.

Le domaine de l'apprentissage automatique est très vaste. On peut cependant le diviser grossièrement en deux parties : d'une part, la classification d'ensembles et la construction automatique de classes; d'autre part, l'induction de règles de grammaire. Par ailleurs, il se pose un choix de méthodes. Certaines reposent sur des techniques statistiques (Quinlan 1993), d'autres sur des techniques symboliques (Michalski & al. 1984).

L'apprentissage automatique a suscité un grand intérêt à un moment mais il n'a pas produit, jusqu'ici, de résultats à la hauteur des espérances. Ceci s'est soldé par des déceptions et peut être par un relatif désintérêt. Les promesses trop nombreuses et le flou de sa définition en ont sans doute été la cause. Nous sommes conscient de ces écueils et nous nous proposons d'appliquer les techniques d'apprentissage à des domaines relativement bien circonscrits tels que l'induction de grammaires et la construction automatique de classes.

5.2.3.1 L'induction de grammaires

Le système syntaxique que nous avons proposé se fonde sur trois grammaires : une pour les groupes verbaux et nominaux, une autre pour les bigrammes, et enfin une pour les dépendances. Un analyseur complet nécessite des règles exhaustives pour chacun de ces niveaux. Un processus d'apprentissage automatique doit donc considérer les trois niveaux de règles.

Pour ce qui concerne le premier niveau, des méthodes symboliques existent pour tirer des règles sur les parties du discours et la détection de groupes (Brill 1995, Ramshaw & Marcus 1995). Elles reposent sur l'existence de corpus annotés. Par ailleurs, il est peut-être possible de construire une grammaire exhaustive des groupes à la main. S. Abney (1996) a proposé une grammaire des groupes nominaux et verbaux de l'anglais qui tient en une page. Il l'applique avec l'algorithme du plus long appariement et obtient de bons résultats.

Pour ce qui concerne les bigrammes et les dépendances, le point est plus complexe. Ceci d'autant plus qu'ils font parfois appel à des traits lexicaux. Collins (1996) a cependant montré qu'il est possible d'inférer des statistiques sur les relations de dépendance en anglais. Il a l'avantage de disposer de corpus étiquetés.

Pour le français, l'inférence de règles serait à tenter et à évaluer. Cependant, on se heurte actuellement au manque de corpus parenthésés facilement accessibles pour mener l'expérience. Au préalable, il sera donc nécessaire d'obtenir ou de construire des corpus parenthésés.

5.2.3.2 La construction de classes

La navigation dans des mondes virtuels repose sur un dictionnaire ontologique construit à la main et tenant compte des relations entre l'espace et la langue. On doit aussi adjoindre à l'agent une capacité de raisonnement sur les mots et les objets du monde. De nouveau, il se pose le problème de créer ou d'enrichir automatiquement la connaissance pour l'étendre à d'autres mondes et à d'autres capacités sans avoir à la reconstruire à chaque fois.

Church & Hanks (1990), Brown et al. (1992), McMahon & Smith (1996) ont décrit des méthodes de construction automatique de classes. Elles leur permettent de tirer une hiérarchie ontologique des mots d'un lexique. La classification différencie automatiquement la nature sémantique des verbes et regroupe, par exemple, les verbes d'ingestion, ceux de destruction et ceux de perception. Ceci est à relier aux classes de verbes, obtenues manuellement, qu'utilise Ulysse et qui doivent distinguer les changements de posture, comme *regarder*, et les changements de lieu. De telles méthodes ne nécessitent que des corpus bruts sans annotation. Elles se fondent sur les proximités de mots en utilisant la mesure d'information mutuelle.

D'autres méthodes utilisent une connaissance syntaxique préalable pour enrichir une base lexicale (Pedersen & Chen 1995, Jacquemin 1995). Ces méthodes concernent l'acquisition de relations syntaxique et sémantique. Elles utilisent une grammaire de surface et tirent parti de phénomènes linguistiques tels que les contraintes de dépendances ou la coordination. Ce type de méthodes permet d'apprendre les propriétés syntaxiques de mots nouveaux, des concepts ou des variantes terminologiques.

Un dernier niveau consiste à établir des relations plus complexes, telles que des liens de synonymie, d'hypermymie, de méronymie, etc. ou d'identifier des rôles thématiques. Dans Ulysse, nous avons exploité de telles fonctions, par exemple avec la décomposition de certains verbes relativement complexes, comme *faire le tour*, dans une séquence de verbes plus simples (Godéreaux et al. 1997). La détermination des relations lexicales se fonde sur les définitions des mots. Bien qu'elle soit encore embryonnaire, l'analyse automatique d'un dictionnaire de définitions permettrait de s'affranchir dans une certaine mesure d'une étude à la main (Vanderwende 1995).

Toutes ces techniques conjuguées devront permettre d'acquérir la terminologie plus facilement sur un grand nombre d'espaces virtuels. On pourra ainsi donner automatiquement à des agents les capacités lexicales correspondantes.

5.2.3.3 Raisonnement et perception

Les interfaces de conversation font appel à l'analyse syntaxique et sémantique mais aussi parfois à un raisonnement élaboré. Il s'agit, pour un agent, de comprendre ce que l'utilisateur veut lui dire pour qu'il puisse répondre ou exécuter une commande.

La syntaxe d'une phrase peut s'analyser avec des méthodes relativement universelles, quels que soient les domaines d'une langue. Sa compréhension, en revanche, requiert à chaque fois une adaptation. Dans la navigation, le point crucial du raisonnement est relié à l'espace et pour l'exécution de l'ordre, à la planification de mouvements. Dans la commande d'un objet ou dans la délégation de tâches, d'autres sémantiques et d'autres raisonnements seraient à mettre en œuvre.

Dans le cas de la navigation dans le cerveau, le raisonnement prend en charge la décomposition de l'action. Ulysse peut planifier un mouvement, par exemple, pour se diriger vers le tronc cérébral. Cependant, la navigation dans le corps humain pourrait faire appel à des commandes telles que *Suis l'artère cervicale*. L'analyse syntaxique de cette phrase est facile, la représentation sémantique au sens de la forme logique, aussi. C'est sur le raisonnement demandé pour exécuter l'action que repose la charge. Dans cet exemple, il fait intervenir un couplage de la langue et d'opérateurs de reconnaissance des formes pour détecter l'objet en jeu. Le raisonnement spatial de la navigation débouche sur l'étude de la perception et du raisonnement tout court dès qu'il s'agit d'être général.

Le raisonnement, surtout lorsqu'il est couplé à la perception, est malheureusement complexe. À chaque application, il serait sans doute à reconstruire en partie. Notre objectif sera de tenter de mieux cerner ce qui peut s'apprendre automatiquement par une connaissance dérivée de corpus, par exemple pour regrouper des verbes de même signification; quelles sont les définitions qu'on peut tirer automatiquement d'un dictionnaire, etc. et enfin, peut-être, comment instruire une machine à l'aide d'explications ou d'exemples en langage ordinaire.

5.2.4 Des études cognitives

Pour l'élaboration et le développement de nos nouveaux projets, nous avons l'intention d'intégrer de façon plus systématique des études cognitives. Dans notre premier prototype, nous n'avons fait qu'effleurer leur examen. Ces études concerneront notamment :

- les relations entre la langue et l'image
- les dénominations d'objets et les préférences dans la description d'actions
- la multimodalité

5.2.4.1 La langue et l'image

Un agent conversationnel dans un monde virtuel fait appel à des représentations et des constructions mentales qui sont des combinaisons visuelles et linguistiques. Michel Denis (1996) remarque que

l'image et le langage constituent deux modes de représentations [...] fortement différenciées, mais dont la coopération est requise dans de nombreuses formes de fonctionnement cognitif.

Cette articulation est particulièrement importante quand il s'agit de naviguer à la voix, et au delà de commander et de déléguer, dans un espace tridimensionnel.

Nous n'aurons sans doute pas les moyens de mener des recherches exhaustives dans le domaine de la perception et de la représentation. Cependant, nous tâcherons d'y contribuer par le biais de l'expérimentation et de la collecte de corpus. Le fait que nous disposions d'un système combinant un monde virtuel et une interface linguistique nous donne un outil pour conduire ces expérimentations. Un tel système peut aider à valider ou à infirmer plus facilement des théories sur l'articulation entre l'image et le langage. Il permet aussi de préciser des intuitions et peut être de formaliser une idée.

Les relations entre les mots et leur mise en correspondance dans l'espace ont été exploitées pour la visualisation et la navigation dans un espace d'informations. Le centre de recherche de Xerox à Palo Alto a notamment produit un certain nombre de ces systèmes. Robertson et al. (1993), plus récemment Card et al. (1996) et Hearst et Karadi (1997) ont utilisé des métaphores iconiques pour structurer l'information d'une base de textes.

La structure prend la forme de « mondes », plus ou moins éloignés suivant la distance entre les concepts qui leur sont attachés. À l'intérieur des mondes se trouvent les documents que les concepts recouvrent. Les algorithmes d'analyse linguistique utilisent les notions de *Vector Space* développées par Salton (1989) pour calculer la proximité entre les documents. D'autres réalisations permettent de structurer les pages des documents en fonction de la page qu'on désire lire et de ses voisines (Figure 28). D'autres permettent de rattacher un document à une hiérarchie de concepts projetée dans l'espace (Figure 29).



Figure 28 Le système WebBook (Card et al. 1996).

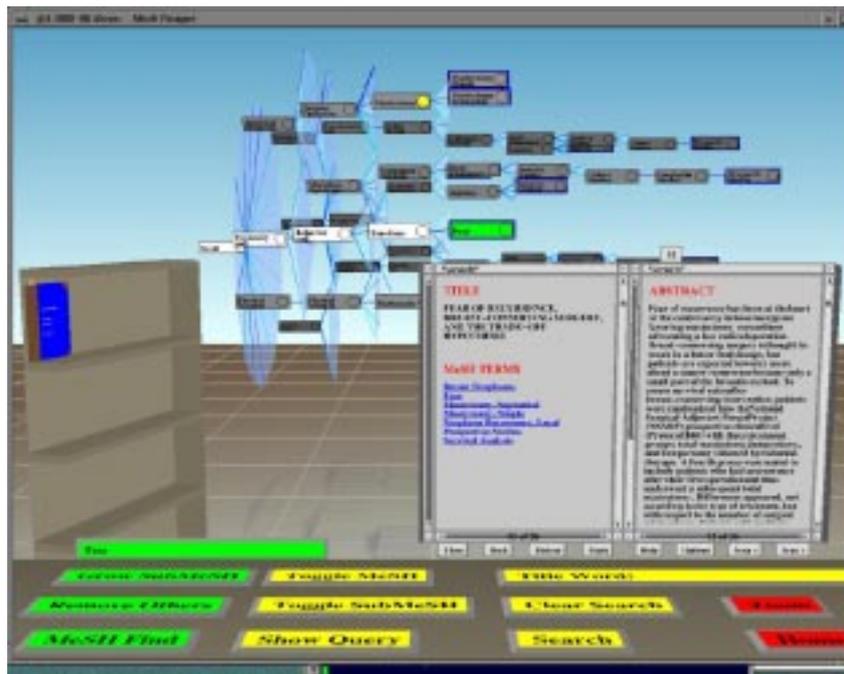


Figure 29. Le système Cat-a-Cone (Hearst & Karadi 1997).

Avec ces interfaces, des masses de données deviennent plus facilement appréhendables. On peut les manipuler : s'en approcher, s'en éloigner, déployer une hiérarchie.

Un agent conversationnel de navigation et de manipulation nous permettra d'examiner comment un utilisateur apprécie les liens entre les distances sémantiques telles qu'il les conçoit et les distances de représentation ou bien entre des requêtes pour manipuler la représentation des concepts dans l'espace et les « mouvements » de sens qui leur correspondent.

5.2.4.2 La dénomination des objets et les préférences dans la description d'actions

La dénomination des objets et l'emploi d'un vocabulaire sont très variables suivant les individus. Furnas et al. (1987) rapportent que pour la description d'un même objet, les utilisateurs utilisent spontanément un même mot avec une probabilité inférieure à 20%. Il est important de collecter de la façon la plus complète possible le lexique attaché aux mondes que nous voulons explorer.

Les problèmes du vocabulaire et de la dénomination sont particulièrement gênants dans les systèmes de travail coopératifs. Chen (1994) a montré qu'on pouvait cependant aboutir à des classes de mots par un algorithme dérivé du Vector Space. Il l'a appliqué pour aider à faire converger une discussion en ligne. Nous pensons que de telles procédures systématiques sont aussi nécessaires pour organiser le vocabulaire relatif à la fois aux dénominations d'objets d'un monde virtuel et aux descriptions d'actions.

Nous amorcerons la construction du lexique par une collecte de vocabulaires appropriés en dépouillant des sources. Puis, pour le compléter, nous procéderons à

des analyses de corpus de description de scènes. Nous essayerons de couvrir le lexique de dénomination d'objets en faisant appel à des utilisateurs sans connaissance particulière du monde et à des utilisateurs au fait du sujet. Par exemple, des latinistes pour le Plan de Rome, des médecins pour le cerveau, des architectes pour les mondes architecturaux et des utilisateurs ordinaires.

La description des actions de navigation et l'utilisation des verbes qui leur sont liées sont aussi sujettes à des préférences. Au delà du simple choix de vocabulaire, il y a l'adéquation entre un verbe de mouvement et une perception psychologique. L'analyse est complexe pour certains verbes comme *contourner* qui ont en général un objet en référence. Le mouvement doit être « naturel » pour que l'utilisateur se sente à l'aise dans sa « peau » virtuelle.

Nous tenterons de préciser comment réaliser la meilleure correspondance entre la description d'une trajectoire ou d'un itinéraire et la planification qui en résulte. Nous tenterons enfin d'établir des mesures de satisfaction d'un utilisateur quant à la transcription de sa description d'un mouvement dans un transport virtuel par l'agent conversationnel.

5.2.4.3 La multimodalité

Nous examinerons l'interaction multimodale et ses rapports avec la cognition en mesurant les préférences des utilisateurs pour les modes linguistiques et gestuels. Pour ceci, nous utiliserons des scénarios de coordination des gestes et de la voix. Ils intégreront des désignations et des descriptions et nous analyserons les corpus d'interaction qui en résulteront.

Nous préciserons en parallèle dans le cadre d'une application de travail coopératif, les apports de la communication multimodale. L'un des problèmes de la communication orale dans une telle application, vient du fait que tous les utilisateurs connectés sont potentiellement récepteurs de tous les messages émis. Il peut en résulter de la confusion. Un utilisateur, lorsqu'il parle, peut s'adresser à un seul utilisateur, ou plusieurs, ou bien vouloir commander des agents. Les agents doivent à tout moment tenter de capter les commandes qui leur sont destinées et les exécuter. Au contraire, ils doivent rester inactifs si la commande ne les concerne pas.

Pour réaliser cette partie, nous ajouterons aux agents un modèle de « conscience » (*awareness*) spatiale et linguistique (Benford & Fahlén 1993). Nous pourrions réaliser l'incorporation de la « conscience » notamment par une modélisation de l'intention. La prise en compte de l'intention des acteurs permettra à un agent de détecter les commandes qui lui sont adressées de manière linguistique ou par geste. Un agent pourra prendre conscience qu'on s'adresse à lui quand l'utilisateur le désignera du doigt, se tournera vers lui, etc. À la suite de quoi, il pourra se déclencher. Dans le cas où les agents sont certains qu'une commande leur est destinée, ils l'exécuteront. Sinon, ils mettront en œuvre une procédure pour clarifier.

Nous intégrerons des scénarios sociaux de coordination pour déterminer s'ils peuvent améliorer la compréhension. Par exemple lors d'un énoncé ambigu est-il préférable de donner des précisions à la voix ou par un geste et dans quel cas? Si plusieurs agents se reconnaissent lors d'un énoncé ou d'une commande, est-il préférable que les agents en question s'annoncent par un énoncé oral, qu'ils se désignent en pointant le doigt vers eux ou qu'ils accourent vers l'utilisateur?

.....

Le développement spectaculaire du réseau Internet a entraîné le partage et la mise en commun d'une connaissance considérable. Il a favorisé la communication immédiate et la transmission du savoir en réduisant le temps et les distances. À l'intérieur d'Internet, il existe maintenant une multitude d'espaces virtuels peuplés d'utilisateurs. Certains auteurs prédisent qu'une vie sociale apparaîtra à travers ce réseau et les espaces qu'il recrée.

Les espaces virtuels d'Internet utilisent parfois des métaphores et des interfaces simulant un monde réel. Elles se fondent sur une interprétation ou une recreation plus exacte de la réalité. La visualisation tridimensionnelle permet de structurer dans l'espace une grande quantité de documents, de reconstruire en temps réel des parties du corps humain à partir de données de capteurs, de rendre compréhensibles des sessions de travail en groupe incorporant la visioconférence.

La puissance de représentation a comme contrepartie une difficulté plus grande dans la maîtrise et l'exploration des univers sous-jacents. Nous avons montré qu'un agent conversationnel réalisant l'interface linguistique entre un utilisateur et un monde virtuel était réalisable et utile.

L'objectif de nos projets futurs est de multiplier de tels agents dotés de capacités linguistiques pour « habiter » les représentations des utilisateurs et des objets. Nous pensons qu'ils leur donneront de nouvelles facultés et permettront une communication plus simple et plus naturelle entre toutes les entités de ces « nouveaux » mondes. On peut aussi concevoir qu'à terme, ils offriront une nouvelle vie à des villes comme Rome, à ses habitants disparus, et qu'ils nous permettront d'imaginer ce qu'ils ont pu être.

6.
Références

Abney S., *Cascaded Finite-State Parsing*, Talk given at Xerox Research Centre, Grenoble, www.sfs.nphil.uni-tuebingen.de/~abney/Papers.html, 1996.

Allen J.F.: *Natural Language Understanding*, Second edition, Benjamin/Cummings, 1995.

Allen, J.F., Schubert, L.K., Ferguson, G., Heeman, P., Hee Hwang, C., Kato, T., Light, M., Martin, N.G., Miller, B.W., Poesio, M. & Traum, D.R., The TRAINS Project: A case study in building a conversational planning agent, *TRAINS Technical Note 94-3*, University of Rochester, New York, September 1994.

Amat J.-L., P. Nugues, J. Blanquart et G. Grenier, Recalage automatique d'objets cartographiques dans une image. Application au recalage radar/carte, *7e congrès de l'AFCEP-INRIA, Reconnaissance des formes et Intelligence artificielle*, 1989, Paris, Tome 2, pp. 803-816.

Andersson, M., Carlsson, C., Hagsand, O. & Ståhl, O., *DIVE, The Distributed Interactive Virtual Environment, Technical Reference*, Swedish Institute of Computer Science, Kista, Suède, March 1994.

Appelt D., Hobbs J., Bear J., Israel D., Kameyama M., & Tyson M., SRI: Description of the JV-FASTUS System used for MUC-5, In *Proc. Fifth Message Understanding Conf. (MUC-5)*, Baltimore, MD, August 1993, Morgan Kaufman.

Ball, G., Ling D., Kurlander D., Miller J., Pugh D., Skelly T., Stankosky A., Thiel D., Van Dantzich M., & Wax T.: Likelike Computer Characters: The Persona Project at Microsoft Research, in *Software Agents*, J. Bradshaw ed., MIT Press, To appear, 1997.

Benford S., Bowers J., Fahlén L, Greenhalgh C., & Snowdon D., User Embodiment in Collaborative Virtual Environments, *CHI'95 Proceedings*, May 1995, Denver, Colorado, 1995.

Benford S & L. Fahlén, A Spatial Model of Interaction in Virtual Environments, *Proceedings of the 3rd European Conference on CSCW*, September 1993,

Bersot O., *Navigation dans les mondes virtuels : Planification d'actions et animation graphique*, Mémoire de Diplôme d'Études Approfondies, Université de Caen, septembre 1996.

Bersot O., P.-O. El Guedj, C. Godéreaux, & P. Nugues, A Conversational Agent to Help Navigation and Collaboration in Virtual Worlds, In: *Proceedings of Collaborative Virtual Environments 1996*, CVE' 96, University of Nottingham, September 1996.

Bescherelle, *L'art de conjuguer*, Hatier, 1980.

Bolt, R. A.: Put That There: Voice and Gesture at the Graphic Interface, *Computer Graphics*, vol. 14(3), pp. 262-270, 1980.

Brill Eric, Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part of Speech Tagging, *Computational Linguistics*, 21(4):543-565, 1995.

Brown P. F., V. J. Della Pietra, P. V. deSouza, J. C. Lai, & R. L. Mercer, Class-Based n-gram Models, of Natural Language, *Computational Linguistics*, 18(4):467-479, 1992.

- Card S. K., G. G. Robertson, & W. York, The WebBook and the Web Forager: An Information Workspace for the World-Wide Web, *CHI'96 Proceedings*, April 1996, Vancouver, Canada, 1996.
- Cazenave F., P.-O. El Guedj, C. Godéreaux & P. Nugues, Un système de dialogue oral guidé pour la génération de comptes rendus médicaux, In: *Actes du 9e congrès de l'AF CET-INRIA, Reconnaissance des formes et Intelligence artificielle*, Vol. 2, Paris, janvier 1994, pp. 79-88.
- Chen H. Collaborative systems: solving the vocabulary problem. *IEEE COMPUTER*, 27(5):58-66, Special Issue on Computer-Supported Cooperative Work (CSCW), May 1994.
- Chen H., P. Hsu, R. Orwig, L. Hoopes, and J. F. Nunamaker. Automatic concept classification of text from electronic meetings. *Communications of the ACM*, 37(10):56-73, October 1994
- Chomsky N.: *Syntactic Structures*, Mouton, 1957. [trad. fr. *Structures syntaxiques*, éd. du Seuil, 1969]
- Chomsky N.: *Cartesian Linguistics*, Harper and Row, 1966. [trad. fr. *La linguistique cartésienne*, éd. du Seuil, 1969]
- Chomsky N., Remarks on nominalization, In: R Jacobs and P. Rosenbaum (eds), *Readings in English Transformational Grammar*, Ginn, 1970.
- Church K.: A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text, *Second Conference on Applied Natural Language Processing*, ACL, 1988.
- Church K. & P. Hanks, Word Association Norms, Mutual Information, and Lexicography, *Computational Linguistics*, 16(1), 22-29, 1990.
- Church K. & Mercer R., Introduction to the Special Issue on Computational Linguistics Using Large Corpora, *Computational Linguistics*, 19(1):1-24, 1993.
- Collins M.J., A New Statistical Parser Based on Bigram Lexical Dependencies, cmp-lg/9605012, *EPrint Archive*, xxx.lanl.gov/cmp-lg, Internet, 1996.
- COTECH, Minutes of the COTECH Workgroup: Virtual and Augmented Environments for CSCW, Department of Computer Science, University of Nottingham, Nottingham, England, 1995.
- Covington M.: Parsing discontinuous constituents in dependency grammars, *Computational Linguistics*, 16(4):234-236, 1990.
- Covington M.: *An Empirically Motivated Reinterpretation of Dependency Grammar*, Research Reports AI-1994-01, University of Georgia, 1994.
- Crangle C. & P. Suppes, *Language and Learning for Robots*, CSLI Publications, 1994.
- de Kleer J., An Assumption-Based TMS. *Artificial Intelligence*, 28:127-163, 1986.
- Denis M., *Rapport scientifique du Groupe Cognition humaine*, LIMSI, Orsay, 1996.
- Diebel K., *Entwicklung eines Dialogsystemes für eine virtuelle Konferenz*, Diplomarbeit der FH Konstanz, August 1994.

- Doyle J., A truth maintenance system, *Artificial Intelligence*, 12:231-271, 1979.
- Ducrot O. & J.-M. Schaeffer, *Nouveau dictionnaire encyclopédique des sciences du langage*, Le Seuil, 1995.
- Earley J.: An efficient context-free parsing algorithm, *Communications of the ACM*, 13:92-102, 1970.
- El Guedj, P.O. & Nugues, P., A chart parser to analyze large medical corpora, In: *Proceedings of the 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Baltimore, pp. 1404-1405, November 1994.
- El Guedj P.-O., *Analyse syntaxique par chart combinant règles de dépendance et règles syntagmatiques*, Mémoire de thèse, Université de Caen, 10 septembre 1996.
- El Guedj P.-O. & P. Nugues, Un analyseur syntaxique combinant plusieurs formalismes au sein d'un chart hiérarchique, *Actes du Colloque RÉCITAL '96*, Dourdan, septembre 1996.
- Engelmore R. & T. Morgan (eds), *Blackboard Systems*, Addison-Wesley, 1988.
- Enjalbert P. & B. Victorri, Du langage au modèle, *Traitement automatique des langues*, 35(1):37-64, 1994.
- Everett S., Wauchope K., Perez M.A.: A Natural Language Interface for Virtual Reality Systems, *Technical Report of the Navy Center for Artificial Intelligence*, US Naval Research Laboratory, Washington DC, 1995.
- Fikes R. & Nilsson N. J. STRIPS: A new approach to the application of theorem proving to problem solving, *Artificial Intelligence*, 2, 189-208, 1971
- Furnas G. W., T. K. Landauer, L. M. Gomez, S. T. Dumais, The vocabulary problem in human-system communication, *Communications of the ACM*, 30(11):964-971, 1987
- Gaizauskas R., T. Wakao, K. Humphreys, H. Cunningham, Y. Wilks, Description of the LaSIE System as Used for MUC-6, in: *Message Understanding Evaluation and Conference: Proceedings of the Sixth ARPA Workshop*, Morgan Kaufmann, 1996.
- Gentner Don & Nielson Jakob: The Anti-Mac Interface, *Communications of the ACM*, Vol. 39(8), August, 1996, pp. 70-82.
- Fradin B., L'approche à deux niveaux en morphologie computationnelle et les développements récents de la morphologie, *Traitement automatique des langues*, 35(2), 1994, pp. 9-48
- Godéreaux C., P.-O. El Guedj, F. Revolta, & P. Nugues, Un agent conversationnel pour naviguer dans les mondes virtuels, *HumanKybernetik*, Band 37, Heft 1, pp. 39-51, 1996.
- Godéreaux, C., Diebel, K., El-Guedj, P.O., Revolta, F. & Nugues, P.: An Interactive Spoken Dialog Interface to Virtual Worlds, in: *Linguistic Concepts and Methods in CSCW*, J.H. Connolly & L. Pemberton eds, Chapter 13, pp. 177-200, Springer Verlag, 1996.
- Godéreaux C., *Un modèle d'agent conversationnel pour naviguer dans un monde virtuel*, Mémoire de thèse, Université de Caen, 7 janvier 1997.

- Godéreaux C., P.-O. El Guedj, F. Revolta, & P. Nugues, Ulysse: An Interactive Spoken Dialogue Interface to Navigate in Virtual Worlds, Lexical, Syntactic, and Semantic Issues, in: *Proc. of the International Conference From Desktop to Webtop: Virtual Environments on the Internet, WWW and Networks*, Bradford, 14-17 April 1997.
- Gong Y. & J.P. Haton, A specialist society for speech understanding, in: *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1988.
- Grosz B. & C. Sidner, Attention, intentions, and the structure of discourse, *Computational Linguistics*, 12:175-204, 1986.
- Grudin, J.: Computer-Supported Cooperative Work: History and Focus, *Computer*, vol. 27(5), pp. 19-26, May 1994.
- Harris Z.S.: *Mathematical Structures of Language*, Wiley, 1968.
- Haton J.P., N. Bouzid, F. Charpillet, M.C. Haton et al., *Le raisonnement en intelligence artificielle : Modèles, techniques et architectures pour les systèmes à bases de connaissance*, InterEditions, 1991.
- Hearst M. A., TextTiling: Segmenting Text into Multi-paragraph Subtopic Passages, *Computational Linguistics*, 23(1):33-64, 1997
- Hearst M. A., C. Karadi, Cat-a-Cone: An Interactive Interface for Specifying Searches and Viewing Retrieval Results using a Large Category Hierarchy, in: *Proc. of 20th Annual International ACM/SIGIR Conference*, Philadelphia, PA, July, 1997.
- Hudson Richard, *Word Grammar*, Blackwell, 1984.
- Huls, C., Bos, E. & Claassen, W.: Automatic Referent Resolution of Deictic and Anaphoric Expressions, *Computational Linguistic*, vol. 21(1), pp. 59-79, 1995.
- Ide & N., Véronis, J. (Eds.) (1995). *The Text Encoding Initiative: Background and Context*. Dordrecht: Kluwer Academic Publishers.
- IBM: IBM VoiceType Developer's toolkit for Windows 95, *Programmer's Reference & API Reference*, 1996.
- Jacquemin C. A Symbolic and Surgical Acquisition of Terms through Variation, In: *Proc. Workshop on New Approaches to Learning for NLP at 14th Int. Joint Conf. on Artificial Intelligence*, 1995
- Jung M., N. Badler, & T. Noma, Animated human agents with motion planning capability for 3d-space postural goals. To Appear in *The Journal of Visualization and Computer Animation*, 1994.
- Kahn P., From the genome to proteome: Looking at a cell's proteins, *Science*, Vol. 270, 20 October 1995, pp. 369-370.
- Karlgren, J., Bretan, I., Frost, N. & Jonsson, L.: Interaction Models, Reference, and Interactivity in *Speech Interfaces to Virtual Environments*, 2nd Eurographics Workshop, Monte Carlo, Darmstadt, Fraunhofer IGD, 1995.
- Koskenniemi, K., *Two-level morphology: a general computational model for word-form recognition and production*, Publication n° 11, Department of General Linguistics, University of Helsinki, 1983.

- Lenoir A., R. Malgouyres, M. Revenu, Fast Computation of the Surface's Normals of a 3-D Discrete Object, *Proceedings Discrete Geometry for Computer Imagery '96*, Springer, 1996.
- Ludwig M., *Manipulation von Gehirnbildern mit gesprochener Sprache*, Diplomarbeit der FH Ravensburg-Weingarten, Oktober 1996.
- Ludwig M., A. Lenoir, & P. Nugues, A Conversational Agent to Navigate into MRI Brain Images, in: *Proc. Interfaces 97*, Montpellier, 28-30 mai, 1997.
- Magerman D.M.: *Natural Language Parsing as Statistical Pattern Recognition*, PhD Dissertation, Stanford University, 1994.
- Marcus M. Santorini B., & Marcinkiewicz M: Building a Large Annotated Corpus of English: the Penn Treebank, *Computational Linguistics*, 19(2):313-330, 1993
- McMahon J. & F. Smith, Improving Statistical Language Model Performance with Automatically Generated Word Hierarchies, *Computational Linguistics*, 22(2):217-247, 1996.
- Mel'cuk I.A.: *Dependency Syntax: Theory and Practice*, State University of New York Press, 1988.
- Merialdo B, Tagging English Text with a Probabilistic Model, *Computational Linguistics*, 20(2):155-171, 1994.
- Miller G. A. WordNet: A Lexical Database of English, *Communication of the ACM*, November 1995, 39-41.
- Michalski R. S., J. G. Carbonell, & T. Mitchell, *Machine Learning: An Artificial Intelligence Approach*, Springer Verlag, 1984.
- Nowak Rachel, Entering the postgenome era, *Science*, Vol. 270 (20), October 1995, pp. 368-371.
- Nugues P., J. Steinmetz, J.-P. Haton, M.-M. Galteau, and G. Siest, Using Artificial Intelligence in 2-D Electrophoresis Experiments, Sixth Meeting of The International Electrophoresis Society, *ELECTROPHORESIS '88*, 4th-7th July 1988, Copenhagen, pp. 336-343, University of Copenhagen, Distributor VCH Verlag,
- Nugues P., J. Steinmetz, J.-P. Haton, M.-M. Galteau, and G. Siest, New Perspectives in Data Bases Query Using Artificial Intelligence, *Two-Dimensional Electrophoresis*, ed. by A.T. Endler and S. Hanash, November 8-11, 1988, Vienne, Autriche, VCH Verlag, pp. 58-62.
- Nugues P., *Interprétation de gels d'électrophorèses bidimensionnelles*, Thèse de doctorat, Université de Nancy, Mars 1989.
- Nugues P., R. Whalen et J.-P. Haton, Apprentissage symbolique et interprétation de gels d'électrophorèses bidimensionnelles, *Actes du Congrès INRIA, Analyse des données, Apprentissage de connaissances symboliques-numériques, Complément*, pp. 31-40, 11-14 septembre 1989.
- Nugues P.-M., Two-dimensional Electrophoresis Image Interpretation, *IEEE Transactions on Biomedical Engineering*, 40(8):760-770, August 1993.

- Nugues, P., Cazenave, F., El Guedj, P.O. & Godéreaux, C.: Un système de dialogue oral guidé pour la génération de comptes rendus médicaux, In: *Actes du 9e congrès de l'AFCEP-INRIA Reconnaissance de Formes et Intelligence artificielle*, Paris, vol. 2, pp. 79-88, janvier 1994
- Nugues, P., P.-O. El Guedj, F. Cazenave, and B. de Ferrière, Issues in the Design of a Voice Man Machine Dialogue System Generating Written Medical Reports, In: *Proceedings of the 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 14, n° 3, 1992, pp. 842-844.
- Nugues P., C. Godéreaux, P.O. El Guedj et F. Cazenave, Un système de dialogue oral homme machine pour la génération de comptes rendus médicaux, *Innovation et technologie en biologie et médecine*, Vol. 14, N° 4, pp. 469-480, septembre 1993.
- Nugues, P., Godéreaux, C., El Guedj P.O. & Cazenave, F.: Question answering in an Oral Dialogue System, In: *Proceedings of the 15th Annual International Conference IEEE/Engineering in Medicine and Biology Society*, Paris, vol. 2, pp. 590-591, 1993.
- Nugues P., C. Godéreaux, P.-O. El Guedj, & F. Revolva, A Conversational Agent to Navigate in Virtual Worlds, in: *Proceedings of the Eleventh Twente Workshop on Language Technology*, Susann LuperFoy, Anton Nijholt, and Gert Veldhuijzen van Zanten eds., pp. 23-33, Universiteit Twente, Enschede, June 1996.
- Nugues P. & J.-P. Haton, Un système multi-experts pour l'interprétation d'images d'électrophorèses, In: *Actes du 7e congrès de l'AFCEP-INRIA, Reconnaissance des formes et Intelligence artificielle*, 1989, Paris, Tome 1, pp. 99-113.
- O'Farrell P. H., High resolution two-dimensional gel electrophoresis of proteins, *J. Biological Chem.*, 250:4007-4021, 1975.
- Pedersen T. & W. Chen, Lexical Acquisition via Constraint Solving, In: *Working Notes of the AAAI Spring Symposium on Representation and Acquisition of Lexical Knowledge*, 1995.
- Pereira F. C. N & D. H. D. Warren: Definite Clause Grammars for Language Analysis - A Survey the Formalism and a Comparison with Augmented Transition Networks. *Artificial Intelligence*, 13: 231-278(1980)
- Pied F., C. Poirier, P. Enjalbert & B. Victorri, From language to model, *Proc of ECAI-96*, 1996.
- Pierrel J.-M. *Dialogue oral homme-machine*, Hermès, 1987.
- Pollard, Carl & Sag, Ivan, *Head-Driven Phrase Structure Grammar*, University of Chicago Press, 1994.
- Quéau P., *Le virtuel: vertus et vertige*, Champ Vallon/INA, Seyssel, 1993
- Quéau P., Les nouvelles dimensions du cyberspace, *Le Monde Informatique*, 664:24, 9 février 1996.
- Quinlan J. R., *C4.5: Programs for Machine Learning*. Morgan Kauffmann, 1993.
- Ramshaw L. & Marcus M.: Text Chunking using Transformation-Based Learning, *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, pp. 82-94, 1995.

- Revolta-Bleau F., *Navigation dans les mondes virtuels : Résolution des références, raisonnement géométrique et exécution d'actions*, Mémoire de Diplôme d'Études Approfondies, Université de Caen, septembre 1995.
- Robertson, C.G., Cord, S.K. and Mackinlay, J.D.: Information visualization using 3D interactive animation, *Communications of the ACM*, 36(4):57-71, 1993.
- Roche E & Schabes Y.: Deterministic part of speech tagging with finite state transducers, *Computational Linguistics*, 21(4):227-253, 1995.
- Sabah G., *L'intelligence artificielle et le langage*, 2e ed, Hermès, 1990.
- Sablayrolles, P.: The Semantics of Motion, *Proceedings of the 7th conference of the EACL*, Dublin, 1995.
- Salton G., *Automatic Text Processing*, Addison Wesley, 1989.
- Searle J., Minds, Brains, and Programs, *The Behavioral and Brain Sciences*, 3(3), 1980.
- Schutte J., Virtual Teaching in Higher Education, Internet, <http://www.csun.edu/sociology/virtexp.htm>, 1997.
- Shieber S. *An Introduction to Unification-Based Approches to Grammar*, Lecture notes, 4, Center for the Study of Language and Information Publications, 1986.
- Stabler Jr. Edward, *The Logical Approach to Syntax*, MIT Press, 1992.
- Sung U. J, K. Wohn, A Concurrency Control Model for the Real-time Interactive Shared Virtual Environment, In: *Proceedings of Collaborative Virtual Environments 1996*, CVE' 96, University of Nottingham, September 1996.
- Tesnière, L.: *Éléments de syntaxe structurale*, Klincksieck, 1959.
- VanBogelen, R. A., Abshire, K. Z., Pertsemlidis, A., Clark, R. L. and Neidhardt, F. C., 1996, Gene-Protein Database of Escherichia coli K-12: Edition 6, In *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology*, Second Edition, (eds., F. C. Neidhardt, R. I. Curtiss, C. A. Gross, J. L. Ingraham, and M. Riley, ASM Press, Washington D.C. (in press)
- Vanderwende L., Ambiguity in the Acquisition of Lexical Information, In: *Working Notes of the AAAI Spring Symposium on Representation and Acquisition of Lexical Knowledge*, 1995
- Véronis J. & Khouri L.: Étiquetage grammatical multilingue : le projet MULTTEXT, *Traitement Automatique des Langues*, vol. 36 n°1-2, pp. 233-248, 1995.
- Wilkins D., K. Myers, J. Lowrance, & L. Wesley, Planning and reacting in uncertain and dynamic environments, *Journal of Experimental and Theoretical AI*, vol. 7(1), pp. 197-227, 1995.
- Wirén M., *Studies in Incremental Natural Language Analysis*, Linköping Studies in Science and Technology, Dissertation N° 292, 1992.
- Younger D.H.: Recognition and parsing of context-free languages in time n^3 , *Information and Control*, 10:189-208, 1967.

1. PRESENTATION GENERALE	7
1.1 GRENOBLE	9
1.2 NANCY	10
1.3 CAEN	12
1.4 LES ETUDIANTS QUE J'AI DIRIGES AU COURS DU PROJET ULYSSE ET LEUR THEME DE TRAVAIL	15
1.5 PRESENTATION DU MEMOIRE	16
2. ARCHITECTURE DE SYSTEMES DE RAISONNEMENT : APPLICATION A L'INTERPRETATION D'IMAGES	19
2.1 PRESENTATION	21
2.2 IDENTIFIER DES PROTEINES SUR DES GELS	23
2.2.1 L'intérêt et l'utilisation de l'électrophorèse bidimensionnelle	23
2.2.2 La comparaison classique et ses problèmes	24
2.2.3 L'identification en utilisant des connaissances préalables	24
2.3 UNE ARCHITECTURE POUR IDENTIFIER DES PROTEINES	26
2.3.1 Les principes	26
2.3.2 Les sources de connaissances	27
2.3.3 Objets et communications	28
2.3.4 Agendas	29
2.4 TRAITEMENT D'IMAGES	30
2.5 IDENTIFICATION DES PROTEINES	31
2.6 VERIFICATION ET INTERPRETATION	32
2.7 APPLICATION AUX APOLIPOPROTEINES DU PLASMA	33
2.8 CONCLUSION	35
2.8.1 Perspectives pour l'électrophorèse	35
2.8.2 L'évolution de l'architecture	36
2.8.2.1 DictaMed	36
2.8.2.2 Le recalage d'images radar	37
3. ANALYSE SYNTAXIQUE	39
3.1 PRESENTATION	41
3.2 POURQUOI UNE ANALYSE SYNTAXIQUE?	41
3.3 REPRESENTER LA SYNTAXE	42
3.4 DEUX PARADIGMES POUR REPRESENTER LA SYNTAXE	43
3.4.1 Les grammaires génératives	43
3.4.2 Les grammaires de dépendance	44
3.5 DES REPRESENTATIONS MOINS PROFONDES	45
3.5.1 Pourquoi une représentation moins profonde?	45
3.5.2 L'étiquetage par les parties du discours	45
3.5.3 La détection des groupes	47
3.6 UN SYSTEME STOCHASTIQUE D'ANALYSE COMBINANT PLUSIEURS PARADIGMES	49
3.6.1 Introduction	49
3.6.2 Correspondance entre le modèle syntagmatique et les grammaires de dépendance	49
3.6.3 Le calcul des relations de dépendance	50
3.6.4 L'analyseur syntaxique	51
3.6.5 Discussion	52
3.7 UN SYSTEME SYMBOLIQUE D'ANALYSE COMBINANT PLUSIEURS PARADIGMES	53
3.7.1 L'analyse syntaxique et l'anglais	53
3.7.2 Les étapes de l'expérimentation	54
3.7.3 La structure d'un chart	55

3.7.4	Le premier système et ses extensions	57
3.7.5	L'architecture du second système	58
3.7.6	Les catégories des segments	59
3.7.7	Le découpage des segments	60
3.7.8	L'élagage	61
3.7.9	Les dépendances	62
3.7.9.1	L'algorithme de Covington	62
3.7.9.2	Un nouvel algorithme	63
3.7.9.3	Résultats	64
3.7.10	Discussion des résultats	64
3.8	CONCLUSION ET PERSPECTIVES	65
4.	AGENTS CONVERSATIONNELS DANS UN MONDE VIRTUEL	67
4.1	PRESENTATION	69
4.2	LE TRAVAIL COOPERATIF	69
4.2.1	Pourquoi un environnement virtuel et comment le piloter	71
4.2.2	Interfaces linguistiques dans un environnement virtuel	73
4.3	UN CORPUS DE DIALOGUES	74
4.3.1	Les conditions de l'expérience	74
4.3.2	L'intérêt d'une interaction orale	75
4.4	INTEGRER LE DIALOGUE ORAL DANS UN MONDE VIRTUEL	76
4.5	ULYSSE	76
4.5.1	L'architecture	76
4.5.2	La reconnaissance vocale	78
4.5.3	L'analyse syntaxique	78
4.5.4	Les actes de dialogue et le traitement sémantique	79
4.5.5	Le raisonnement géométrique et la résolution des références	81
4.5.6	La gestion du dialogue	84
4.5.7	La planification et la gestion des actions	85
4.5.8	Un voyage avec Ulysse	87
4.6	L'ADAPTATION D'ULYSSE A LA NAVIGATION DANS LE CERVEAU	91
4.7	CONCLUSIONS ET PERSPECTIVES	93
5.	CONCLUSION	95
5.1	CONTRIBUTIONS ET PROBLEMATIQUE	97
5.1.1	Systèmes de raisonnement pour l'interprétation d'images	97
5.1.2	L'analyse syntaxique	98
5.1.3	L'interaction linguistique dans un monde virtuel	99
5.2	RECHERCHES FUTURES	100
5.2.1	Les enjeux et les objectifs	100
5.2.1.1	L'Internet	100
5.2.1.2	Les domaines de travail	100
5.2.2	Vers des agents conversationnels aux capacités plus étendues	101
5.2.2.1	La langue pour explorer des mondes virtuels	101
5.2.2.2	La langue pour commander	104
5.2.2.3	La langue pour déléguer	106
5.2.3	L'acquisition automatique de connaissances linguistiques	108
5.2.3.1	L'induction de grammaires	108
5.2.3.2	La construction de classes	109
5.2.3.3	Raisonnement et perception	109
5.2.4	Des études cognitives	110
5.2.4.1	La langue et l'âge	110
5.2.4.2	La dénomination des objets et les préférences dans la description d'actions	112
5.2.4.3	La multimodalité	113

6. REFERENCES117