

Architectures logicielles en reconnaissance : parallélisme et modularité

Jean Caelen

Laboratoire CLIPS-IMAG
Domaine universitaire, BP 53
38041 Grenoble Cedex 9

Tél: 76.51.46.27 - Fax: 76.44.66.75
email : Jean.Caelen@imag.fr

Résumé

Cet article propose une architecture de système de compréhension de la parole inspirée d'un modèle cognitif modulariste à base d'une architecture distribuée multi-agents. Compte tenu de la latéralisation des processus cérébraux dans les activités langagières, il est permis d'imaginer une seconde voie rapide reposant sur l'analyse prosodique dont la finalité est de fournir des points d'ancrages pour les processus analytiques *a priori* plus lents de la première voie. Ce paradigme relance l'intérêt des architectures logicielles pour la reconnaissance de la parole et pose des questions sur le contrôle des échanges entre agents pour une meilleure adaptabilité et une meilleure robustesse d'ensemble du système.

1. Introduction

Les systèmes de compréhension automatique de la parole intégrant l'ensemble des traitements acoustiques et linguistiques sont apparus vers 1970, date qui correspond avec une intensification des recherches en IA (Intelligence Artificielle). A cette époque, les systèmes tentent d'utiliser toutes les sources de connaissances possibles (phonétique, lexicale, syntaxe-sémantique, pragmatique, prosodie) en les faisant intervenir au moment le plus opportun dans le processus général de décodage. Le problème de l'organisation des systèmes se pose donc clairement, à la fois pour représenter les connaissances utiles et pour planifier ou gérer leur utilisation. Il faut remarquer que ce problème est rendu plus difficile du fait que les connaissances ne sont pas, *a priori*, homogènes entre elles et qu'elles interviennent à des niveaux différents. Des architectures logicielles sophistiquées ont été proposées qui ont posé le problème fondamental de la stratégie d'interaction entre les différentes sources de connaissances. Ce problème n'est pas encore bien résolu de nos jours, mais de nouvelles voies de recherche apparaissent à travers les architectures distribuées qui permettent un certain parallélisme des processus qui peut rendre caduques les stratégies de planification séquentielle utilisées jusque là. Avant, d'étudier en détail les possibilités offertes par ces architectures, nous allons examiner les propriétés des architectures les plus caractéristiques

devenues maintenant classiques. Sur la figure 1 sont présentés les différents types d'organisation de connaissances dans les systèmes de reconnaissance automatique de la parole [Reddy, 1975] à l'époque des recherches les plus intenses sur cette question (1970-1980). Pour les comparer entre elles, nous définissons les critères suivants :

- Ctl = le contrôle,
- F = la fiabilité
- P = la prédictibilité,
- C = la complexité,
- M = la maintenance,
- TR = le temps de réponse,
- Perf = les performances,
- Fonc = les fonctionnalités.

Dans les systèmes hiérarchiques ascendants (fig. 1.a) l'organisation est très simple : les données acoustiques se propagent du niveau bas (acoustico-phonétique) aux niveaux supérieurs (linguistiques) par abstraction progressive des hypothèses jusqu'à ce que le sens de la phrase soit obtenu. Ce type de système est relativement inefficace parce que pour assurer une certaine fiabilité des informations transmises il faut transférer une grande quantité d'hypothèses d'un niveau à l'autre (stratégie dite en "largeur"), ce qui multiplie inutilement la combinatoire de recherche en faisceau et ne fournit aucune assurance sur la présence de la bonne hypothèse dans la liste propagée. Les processus sont séquentiels ou au mieux en "pipe-line". Par contre, la mise en œuvre de tels systèmes ne pose pas de problème de génie logiciel particulier.

(a) Système hiérarchique ascendant = Organisation taylorienne

- Ctl = gestion "pipe-line" de processus, contrôle simple par les données
- F = repose sur le module le plus fragile (et les niveaux bas)
- P = comportement entièrement prédictible
- C = interchangeabilité des modules si les structures de données sont bien définies
- M = maintenance ramenée à la maintenance des modules de base
- TR = traitement séquentiel donc long
- Perf = recherche en largeur (coûteux)
- Fonc = pas de retour arrière possible en cas d'erreur d'hypothèses, pas d'adaptation globale, pas de guidage sur les "attentes"

Dans les systèmes hiérarchiques descendants ou génératifs (fig. 1.b) la méthode utilisée par chaque niveau est l'analyse par synthèse. En partant du haut, chaque niveau propose des hypothèses à vérifier par le niveau immédiatement inférieur. A son tour, ce niveau, après une série de vérifications sur ses propres connaissances ou en communiquant les hypothèses aux niveaux qui lui sont inférieurs s'il manque d'informations, donne les scores des hypothèses proposées au niveau supérieur. Ainsi une décision peut-être prise à un niveau quelconque généralement sous la forme d'une réfutation puisque les propositions sont faites au niveau supérieur. Ce type de système est relativement répandu à cause de la simplicité à exprimer les connaissances linguistiques de manière générative. Cependant, les limites de ce type de système sont bornées par les limites des représentations linguistiques explicitées dans le système qui interdisent à leur tour pratiquement toute prise en compte des inattendus.

(b) Système hiérarchique descendant = Organisation de marché (offres-demandes)

- Ctl = gestion "pipe-line" d'hypothèses
- F = repose sur la puissance du module le plus haut
- P = comportement entièrement prédictible
- C = raisonnement sur hypothèses
- M = maintenance plus coûteuse que pour le type (a)
- TR = traitement séquentiel mais moins long que le type (a)
- Perf = recherche en profondeur d'abord (risque d'impasse)
- Fonc = retour arrière, pas d'adaptation globale, guidage par les "attentes", décision distribuée, limitation par la puissance du générateur d'hypothèses de chaque niveau

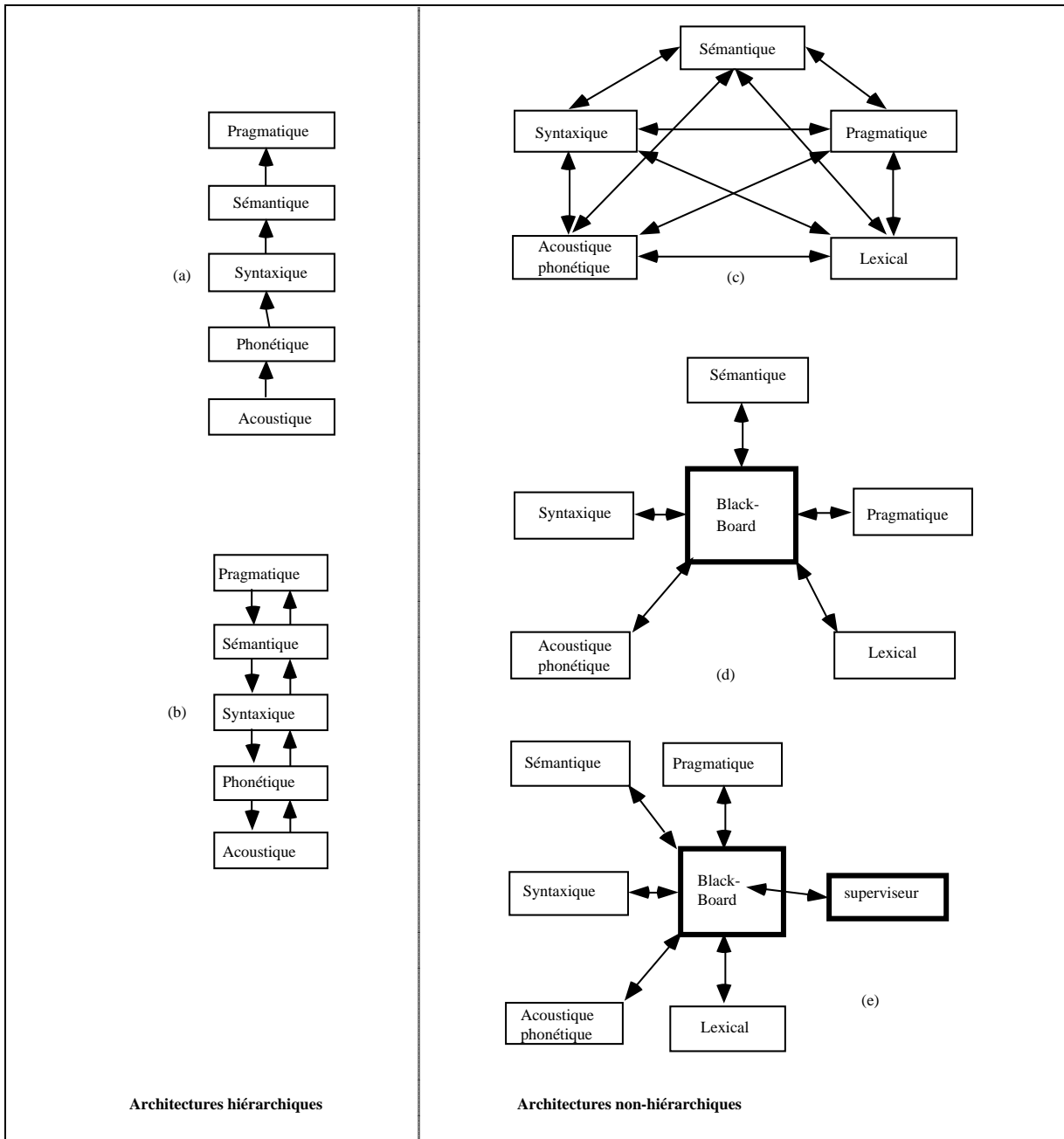


Fig. 1 : Les différents types d'organisation de sources des connaissances dans un système de reconnaissance automatique de la parole. (a) hiérarchique, (b) génératif, (c) non hiérarchique, (d) structure de Black-Board (accès asynchrone), (E) structure de Black-Board avec superviseur (accès synchrone). Les flèches indiquent le sens des échanges d'informations entre les modules.

Les systèmes du type “non hiérarchique” offrent des processus de coopération entre toutes les sources de connaissance ce qui évite les défauts de propagation des hypothèses des systèmes hiérarchiques (fig. 1.c). Ce type d'architecture, dite multi-agents, pose évidemment le problème du contrôle du flux d'informations et de communication des agents.

Une classe importante de tels systèmes est constituée par les systèmes à structure de Black-Board (BB) (Hearsay II par exemple [Lowerre 79]) organisés autour du paradigme du "comité d'experts" (fig. 1.d). Chaque expert peut envoyer des hypothèses à un autre expert à travers le Black-Board qui est une mémoire commune évolutive et qui présente à chaque instant l'état de la situation. Toutes les informations sur les résultats de la reconnaissance à un instant donné, sont regroupées dans le BB dont la structure peut refléter ou non les différents niveaux (phonétique, lexical, syntaxique....). Les experts peuvent lire et écrire de manière anonyme dans un ordre quelconque (stratégie opportuniste) dans le BB dès qu'ils ont quelque chose “à dire”, c'est-à-dire qu'ils peuvent ajouter une information dès qu'ils peuvent la construire au vu des données du BB ou au contraire en supprimer une si elle n'est pas compatible avec les connaissances qu'ils ont.

Les avantages de ce type de structure BB sont :

- 1) de minimiser la duplication des informations,
- 2) que les résultats fournis par un experts sont accessibles à tout moment et sans ordre hiérarchique par les autres experts,
- 3) qu'on peut définir les liens entre les hypothèses des différents niveaux et donc réduire les calculs en cas de retour en arrière ("backtracking"),
- 4) que les étapes de la reconnaissance étant représentés sous forme de graphe, il est possible de poursuivre une stratégie plus générale et d'invoquer des experts en des points où on localise des informations intéressantes ou, au contraire, des difficultés particulières.

(d) Société d'experts = Organisation de séances de “brain-storming”

- Ctl = auto-contrôle, stratégie opportuniste de chaque expert,
- F = un expert peut détruire une info. utile à un autre agent,
- P = comportement non prédictible,
- C = simplifiée par hiérarchisation du BB, complexité individuelle des modules
- M = en apparence facile si les experts n'échangent que des résultats et non des états sur leur raisonnement
 - TR = non-contrôlable (remise en question sans fin des hypothèses), semi-parallèle
 - Perf = îlots de confiance, sens de parcours quelconque (ascendant-descendant), réduction de la quantité de données échangées, focalisation locale sur un problème résistant,
 - Fonc = focalisation des ressources sur les problèmes difficiles, souplesse totale du raisonnement

L'inconvénient de la structure BB est qu'un expert peut détruire une situation ou des connaissances qu'il a jugé sans intérêt alors qu'un autre expert aurait eu besoin de ces connaissances pour son raisonnement. Un autre inconvénient majeur est que l'on n'est pas assuré de la convergence du processus, le comportement de l'ensemble du système n'étant pas prédictible.

Pour éliminer ce dernier inconvénient, un superviseur peut être ajouté à la structure de BB pour la contrôler et planifier les actions des experts. Ce superviseur doit posséder une méta-stratégie de contrôle pour guider le système de compréhension dans son ensemble en évitant les conflits entre les différents experts. Il doit donc avoir une compétence non plus sur le problème de compréhension ou de reconnaissance eux-mêmes mais sur le comportement des experts individuellement et lors de leurs interactions. Il doit de ce fait connaître leur mode de raisonnement. Cette structure BBS (Black-Board Supervisé) est présentée sur la fig. 1.e : le superviseur a ici deux fonctions, (a) gérer les flots de données dans le BB et (b) ordonnancer les actions des experts. Ainsi le problème de la stratégie est de la compétence du superviseur : cela renvoie le problème à un autre niveau et pose celui de savoir quelle est la meilleure méta-stratégie possible pour le superviseur. Une autre question est de définir la granularité des experts, leurs fonctions et compétences, et de modéliser leur raisonnement pour le rendre visible aux autres experts via le superviseur. Le système DIRA [Caelen 90] est un bon exemple de système de compréhension de la parole de type BBS.

(e) Expertise planifiée = Organisation supervisée d'une salle de classe

- Ctl = contrôle des tours d'intervention, des conflits, planification des tâches
- F = repose sur le contrôleur de plans
- P = prédictibilité assurée par le superviseur
- C = gestion des stratégies très complexe
- M = maintenance difficile car tous les modules interfèrent
- TR = séquentiel ou parallèle donc très variable selon les problèmes posés
- Perf = recherche en profondeur guidée par les buts
- Fonc = retour arrière possible, adaptation globale, guidage par les "attentes", décision centralisée, auto-connaissance du système sur lui-même, apprentissage possible

Devant les échecs relatifs des systèmes BB, on a cherché de 1980 à 1990 des architectures plus simples et plus homogènes, fondées comme le système HWIM, sur une représentation unique par réseau en homogénéisant toutes les sources de connaissances — on considère en effet que le concept de *séquence* est apte à rassembler les décompositions de mots en phonèmes aussi bien que de phrases en mots. Cette représentation convient particulièrement bien aux modèles markoviens généralisés. Mais, depuis l'apparition de techniques mixtes, notamment l'utilisation de réseaux de neurones pour la classification avant ou pendant la reconnaissance, ce concept unificateur de séquence n'est plus aussi fécond pour l'intégration de plusieurs techniques. Aussi voit-on depuis quelques années un regain d'intérêt pour les architectures logicielles dans les systèmes de reconnaissance. Ce regain d'intérêt correspond également aux récentes avancées de l'IA en matière de systèmes multi-agents qui fondent le domaine de l'IAD (Intelligence Artificielle Distribuée), c'est-à-dire aux systèmes de la fig. 1.c que nous avons abandonnés pour un temps.

Une des particularités des systèmes multi-agents est le fait qu'ils peuvent être distribués et autonomes, coopérer pour résoudre un même problème, ces deux propriétés rendant leur fonctionnement parallélisable et concurrent. Ainsi ils offrent quelques attraits pour traiter la robustesse d'une part — par coopération de traitements complémentaires en parallèle — et l'efficacité d'autre part — par intégration de traitements concurrents. Ainsi se retrouve-t-on de nouveau en face des questions du choix des modules, de la définition de leur granularité, du contrôle du système, de la coopération-concurrence des agents, de l'organisation "sociale" de leur travail, etc. On distingue actuellement en IAD deux grandes classes d'agents :

(a) les agents réactifs, qui répondent de manière réactive, qui sont à “grain fin” et qui ont des capacités de raisonnement limitées,

(b) les agents cognitifs à “gros grain” qui ressemblent beaucoup aux experts et qui ont des capacités de raisonnement.

Le principe général des systèmes multi-agents est qu'ils travaillent sans contrôle externe apparent, à la manière des sociétés d'animaux ou humaines (ou de modèles biologiques ou neuro-biologiques). La notion de comportement émergent est essentielle dans ces systèmes. Certains agents peuvent se reproduire ou se dupliquer pour augmenter le parallélisme ou “réunir des forces” supplémentaires pour résoudre un problème. Bref les sources d'inspiration et les métaphores sont nombreuses dans le domaine de l'IAD ce qui à la fois nuit quelque peu à sa crédibilité mais qui d'un autre côté lui donne une richesse qu'il ne faut pas rejeter *a priori*.

(c) Multi-agents = Organisation par répartition du travail et coopération

- Ctl = émergentiste sous contraintes, les conflits se résolvent spontanément
- F = repose sur le bon choix des agents et leur compétence
- P = comportement non prédictible
- C = interchangeabilité théorique des agents
- M = difficulté dans la gestion des messages
- TR = temps de stabilisation parfois long
- Perf = parallélisme
- Fonc = décision émergente, adaptation globale, guidage à la fois sur les "attentes" et les entrées

Pour explorer ce domaine et ses possibilités en traitement de la parole, nous avons tenté de mettre en œuvre un système de compréhension automatique de la parole (MICRO) fondé sur les concepts de l'IAD appliqués à la neuro-psychologie. Nous partons du constat que l'humain a des facultés d'adaptation aux phénomènes inattendus et spontanés que n'a pas la machine : la plupart des systèmes actuels de reconnaissance de la parole sont construits sur un principe hiérarchique découpant les fonctions de classification, de filtrage et d'interprétation, en blocs indépendants et figés — ainsi en est-il des systèmes à bases de connaissances hétérogènes mais aussi des systèmes statistiques ou des systèmes auto-organiseurs. Il est bon de constater qu'en dépit de leurs bonnes performances d'ensemble, ces systèmes restent incapables d'affronter les situations inattendues ou de s'adapter à des conditions ou locuteurs nouveaux, facultés pour lesquelles l'être humain est particulièrement efficace.

Il peut paraître intéressant de ce point de vue de s'inspirer des théories récentes en neuro-psychologie pour définir l'architecture d'un système ayant des capacités d'adaptation similaires à celles de l'humain. Parmi les théories disponibles, l'hypothèse modulariste fournit un cadre possible de simulation sous forme distribuée à l'aide d'agents "réactifs" et "cognitifs". Détaillons maintenant ces points.

2. Approche cognitive

Pour concevoir un système de compréhension automatique de la parole, trois grandes caractéristiques de la cognition semblent intéressantes : la modularité, l'interactivité et la co-existence de deux voies d'analyses (analytique et globale).

2.1. Modularité

De nombreux travaux en psychologie cognitive présentent le cerveau comme un ensemble de modules coopérant entre eux tout en opérant chacun dans leur propre domaine de compétence [Minsky 85][Fodor 83][Marr 82]. De même, la neuropsychologie progressant vers une description de plus en plus précise de l'anatomie fonctionnelle du cerveau humain accrédite elle aussi la théorie modulariste tout en en révisant certains aspects, notamment la notion d'irrépressibilité des modules périphériques de Fodor. Actuellement, de nombreux chercheurs présentent sous forme distribuée la structure fonctionnelle des activités cognitives impliquées dans la vision [Kosslyn 80] ou la parole [Morton 84][Altman 90:12-14]. La question de l'indépendance et/ou de l'autonomie de ces modules n'est pas tranchée, non plus que la question de leur granularité. Fodor et Minsky (et bien d'autres) s'opposent sur ce point. Par ailleurs il semble également que les modules soient ramifiés en "couches" — ces couches pouvant être associées physiologiquement aux colonnes corticales qui sont des paquets de neurones connectés entre eux.

2.2. *Interactivité*

La compréhension du langage naturel par l'humain semble être un processus hautement intégré [Mehler 90] dans lequel il semble bien difficile de démêler des niveaux de traitement, —par exemple sémantique et syntaxe s'enchevêtrent intimement. Il est clair que stratégie ascendante et stratégie descendante sont des expressions vides de sens en cognition humaine. Ainsi, la psycholinguistique nous propose actuellement des modèles de reconnaissance de la parole essentiellement interactifs [Mc Clelland et Elman 86][Marslen-Wilson 87]. S'il est nécessaire en simulation par ordinateur de définir des hiérarchies qualitatives de traitements, cela nous entraîne à supposer un processus de diffusion des informations et des mécanismes de parallélisme pour les activités de computation des agents. Nous définissons donc sous le terme d'*interactivité* la production et la prise en compte simultanées des informations ascendantes (stratégie *bottom-up*) et descendantes (stratégie *top-down*). L'interactivité, vue comme une conséquence de la modularité, implique que tous les agents peuvent opérer en parallèle en échangeant leurs informations par des processus diffusants. De cette interactivité émerge un artefact d'organisation qui donne peut-être l'illusion qu'il y a un niveau de contrôle dans le système.

2.3. *Deux voies d'analyse*

Sperry a montré dès les années soixante une différenciation fonctionnelle entre l'hémisphère droit du cerveau, chargé des processus cognitifs non verbaux, et l'hémisphère gauche siège de l'analyse du langage [Sperry 61]. La réalité est sûrement moins simple, et la controverse sur cette latéralisation est loin d'être close. Néanmoins, les travaux en neuropsychologie récents s'accordent sur un point : la co-existence d'une voie d'analyse détaillée (analytique) et d'une voie de traitement plus global (holistique), sans présumer de leur localisation physiologique [Gardner 81]. Ainsi la parole dans ses aspects verbal et acoustique subirait de tels traitements différenciés. Ayant par ailleurs observé empiriquement la nécessité d'une double analyse dans le système DIRA — pour éviter les "retours-arrière" induits par un raisonnement séquentiel en profondeur d'abord et incompatibles avec la gestion temps-réel du signal parole — nous proposons de prendre en compte ces deux voies de la manière suivante :

- une voie analytique parcourue par les processus de décodage acoustico-phonétique et linguistique,
- une voie de pré-marquage rapide reposant sur l'analyse de la prosodie, pour isoler des îlots homogènes et poser des points d'ancrage.

2.4. Principes pour la simulation

Passant trop rapidement sur cette discussion qu'il faudrait approfondir, nous pouvons retenir en première approximation pour la simulation d'un système de compréhension de la parole, les principes suivants :

(a) la couche de "perception périphérique" est composée d'agents réactifs, qui à une stimulation produisent une réponse, cette réponse est transmise à la couche intermédiaire,

(b) la couche d'analyse intermédiaire est composée d'agents cognitifs rassemblés en modules et attachés à des fonctions ; ces agents gardent la maîtrise de leur propre contrôle,

(c) la couche d'intégration est composée d'agents réactifs ou cognitifs qui met en communication des ensembles de modules de la couche intermédiaire et régule leurs interactions.

(d) les couches sont divisées en deux voies, qui pour la couche intermédiaire ont deux fonctions différentes. Ces deux voies fonctionnent en parallèle.

Dans des simulations antérieures et en particulier à travers l'expérience acquise à travers le système DIRA [Caelen 90] on a montré les limites adaptatives d'un modèle modulaire où les modules n'ont pas la maîtrise de leur contrôle. Il semble donc qu'à ce stade de la conception du système, il soit judicieux de laisser un contrôle lâche de type régulation plutôt qu'un contrôle strict de type planification. Cette constatation se justifie aussi à la suite de l'expérience des systèmes BB (Hearsay II par exemple) dans lesquels l'absence de tout contrôle conduisait à des dysfonctionnements fâcheux. Il reste cependant à définir clairement ce qu'est cette régulation : compétition, coopération des agents ou bien quelque autre stratégie ?

3. Description du système MICRO

Le système *MICRO* est développé dans un environnement générique (MAPS=Multi-Agent Problem Solver) dédié à la conception de systèmes multi-agents [Baujard 90].

3.1. MAPS

La conception de MAPS [Baujard 90] repose sur la distinction entre deux classes de connaissances :

- *les connaissances descriptives* sur les éléments du problème (données, hypothèses, résultats),
- *les connaissances opératoires et décisionnelles* (règles et méta-règles), définissant des opérations et un mécanisme de raisonnement sur les éléments du problème.

MAPS associe un type d'agent à chaque classe de connaissance, respectivement les KS (Knowledge Server) et les KP (Knowledge Processor). Les agents communiquent par envoi de messages et réagissent ainsi à des événements extérieurs. Les agents KS communiquent usuellement avec des agents KP (soumission de données et problèmes), et réciproquement les agents KP avec des agents KS (transmission et demande d'informations). MAPS permet également une exécution en parallèle des processus attachés aux agents ; deux modes de synchronisation (synchrone et asynchrone) entre les agents ont été introduits pour permettre la définition de différents comportements (agents indépendants, coordonnés ou concurrents). Les communications en mode synchrone correspondent à des relations *client-serveur*, tandis qu'en mode asynchrone, les communications sont de type *boîte aux lettres*. Enfin, MAPS

permet la création d'architectures multi-couches par l'introduction de méta-agents, chargés du contrôle respectif des tâches et des connaissances.

3.2 Architecture du système MICRO

Le système *MICRO* (fig. 2) constitue un réseau d'agents MAPS. On remarque sur la fig. 2 la structure en trois couches du système, (a) la couche périphérique constituée de deux modules acoustiques simulant le système auditif, (b) la couche intermédiaire constituée de trois modules, deux modules analytiques et un module holistique, simulant classiquement les niveaux phonétique, linguistique et prosodique, et (c) une couche d'intégration et de régulation. La régulation est assurée par des agents Méta-KP et l'intégration est assurée par des agents Méta-KS. Ces agents assurent aussi la communication entre les voies, les couches et l'extérieur. On distingue les deux voies de traitement holistique et analytique symboliquement placées à droite et à gauche.

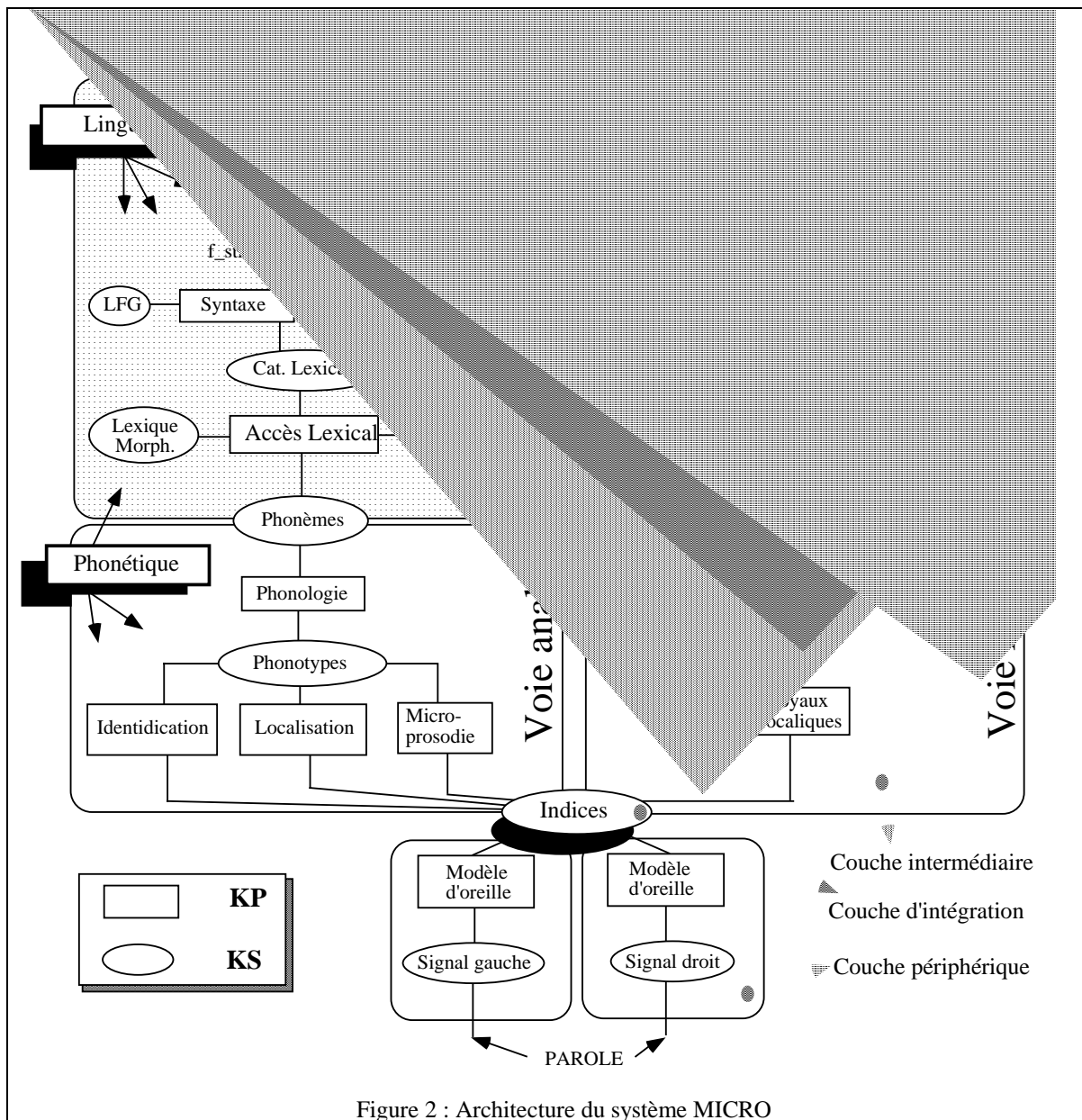


Figure 2 : Architecture du système MICRO

- *Couche périphérique*

Ce groupe d'agents correspond aux traitements du système auditif humain. Il est simulé à l'aide d'un modèle mathématique d'oreille [Caelen 79] et d'un détecteur d'indices acoustiques comportant plusieurs étages de traitement. Tous les agents sont réactifs, ils fournissent une sortie à une stimulation. Ces sorties arrivent dans le Méta-KP "indices" qui joue le rôle de mémoire tampon circulaire et redistribue les données à la couche intermédiaire.

- *Couche intermédiaire*

La couche intermédiaire comprend trois modules. Chaque agent de ces modules reçoit des informations des agents avec lesquels il est connecté que ce soit des modules placés plus bas ou plus haut. Ainsi on perd la notion de stratégie ascendante ou descendante des systèmes d'IA classiques au profit d'une stratégie diffusante dans laquelle joue la synchronisation des informations disponibles.

Module "acoustico-phonétique"

Ce groupe d'agents rassemble les traitements bas-niveau de la voie analytique. Les agents qui le composent produisent des hypothèses phonétiques à partir d'une représentation en indices de la parole et des contraintes phonétiques fournies par le module linguistique. On ne fait pour le moment aucune hypothèse sur la réalisation de ces agents (HMM, NN ou modèles mixtes).

Module "linguistique"

Ce groupe d'agents à la charge de produire la représentation conceptuelle de la phrase, en intégrant les hypothèses phonétiques du groupe acoustico-phonétique, les contraintes de haut-niveau provenant de l'extérieur et les informations prosodiques fournies par la voie holistique (essentiellement de frontières de mots). L'analyse linguistique est mise en oeuvre à l'aide d'un mécanisme de coopération entre les agents syntaxique (grammaire LFG) et sémantique qui travaillent concurremment [Antoine 93]. Comme le montre la figure 3, cette coopération est réalisée par l'intermédiaire d'une fusion de leur productions concurrentes, qu'elles soient ascendantes (structures hypothétiques de la phrase) ou descendantes (contraintes sur l'analyse lexicale).

Cette équivalence des analyses syntaxique et sémantique est attestée à l'heure actuelle par de nombreux travaux présentant l'analyse linguistique comme un processus interactif et intégré [Rastier 87,p. 12]. Ils rompent ainsi avec la tripartition classique plaidant pour une succession ascendante des traitements syntaxique, sémantique et finalement pragmatique [Morris 46].



Module "prosodie"

Ce groupe d'agents constitue l'ensemble de la voie holistique. Il effectue des traitements rapides basés sur l'analyse prosodique du signal de parole, afin de fournir des informations (démarcation de groupes, intonation, etc.) qui serviront de points d'ancrages au module linguistique et pragmatique pour leurs analyses combinatoires. Il définit donc des îlots de confiance *a priori*. (d'où l'exigence que la détection de ces îlots soit fiable).

- *Couche d'intégration (et de régulation)*

Les agents KP de chaque module de la couche intermédiaire construisent, dans leur domaine de compétence, des solutions partielles qui sont regroupées dans un (ou plusieurs) agents Méta-KS de la couche d'intégration. Le problème consiste donc à faire *émerger* les solutions globales du module par confrontation de ses diverses solutions partielles puis à les diffuser vers les autres modules qui ont fait de même pendant le même temps en ignorant les résultats des modules voisins. On pourrait alors aisément imaginer un processus global de stabilisation dont l'état émergent produirait la solution finale en mettant en relation tous les modules. Mais il resterait une décision finale à prendre. La question fondamentale reste donc de savoir si un tel système doit avoir un processus de décision autoritaire ou collective et si l'émergence doit être naturelle ou canalisée. C'est donc de l'utilité des agents Méta-KP que nous allons discuter maintenant pour réguler ce processus : suffit-il comme nous l'avons annoncé dans le § 2.4, d'un contrôle lâche des modules que nous avons appelé régulation ? Si oui quel type de régulation ?

Avant d'expérimenter un véritable système, nous avons tenté une simulation avec un contrôle en deux passes :

- un processus d'unification dans chaque groupe d'agents (qui produit une émergence de solutions partielles),
- la diffusion, avec un léger retard temporel, de ces solutions partielles aux autres groupes d'agents qui permet à ces derniers d'intégrer ces nouvelles données comme contraintes sur leur solution partielle pour le prochain segment de parole.

Ainsi, c'est la prise en compte en parallèle d'informations élaborées par divers processus de nature différente qui autorise un comportement réellement adaptatif [Lautrey 90], faculté primordiale en compréhension de la parole. En effet, l'existence de points de vue différents est essentielle pour la remise en cause d'hypothèses éventuellement erronées.

Nous allons maintenant décrire à titre d'exemple la réalisation du module prosodique (le cadre de cet article ne permettant pas de tout décrire) puis nous reviendrons sur le problème de l'intégration et de la régulation.

4. Le module prosodique

Le module prosodique consitue à lui seul la deuxième voie d'analyse holistique fonctionnant en parallèle avec les modules de la première voie. Son rôle est d'effectuer des repérages rapides sur la surface de l'énoncé pour le découper en éléments plus courts, mots, groupes, syntagmes, etc., d'une part, et de marquer les éléments accentués d'autre part. Il fournit donc des informations qui doivent permettre de synchroniser les traitements du module linguistique sur des îlots de confiance (KP démarcation de mots et de groupe) et de

fournir des informations pour les actes de dialogue (KP intonation) : il fonctionne donc essentiellement en prédiction sauf dans le cas où les attentes provenant des actes de dialogue sont très contraignantes (attentes de réponse de confirmation par exemple). Ce module est constitué d'agents chargés de ces différentes fonctions que nous décrivons quelque peu ci-après.

4.1. KP Noyaux vocaliques.

La prosodie de phrase étant portée par les syllabes, il y a lieu tout d'abord de les détecter. Or à ce stade du décodage, les informations de syllabe ne sont pas encore connues puisque les phonèmes n'ont pas été encore reconnus (l'agent qui les identifie est dans la voie gauche, il est plus lent que les agents de la voie droite). On ne peut donc détecter à ce niveau que des pseudo-syllabes si l'on veut ne pas retarder les traitements de la voie droite, en affectant par exemple cette fonction à un agent KP. Cet agent détecte d'une part les noyaux vocaliques à partir des courbes d'énergie et de fréquence fondamentale fournies par les agents de bas niveau, et d'autre part met à la disposition des agents de plus haut niveau un ensemble de paramètres utiles pour le repérage des groupes prosodiques. Ce sont :

- l'énergie moyenne sur chaque noyau,
- la fréquence fondamentale moyenne,
- la durée de chaque noyau,
- l'accélération du débit d'élocution,
- la durée de pause éventuelle précédant le noyau.

La détection des noyaux vocaliques est un traitement-clef puisque tous les autres agents s'appuient sur cette information pour fournir la leur. Des contraintes de rapidité de calcul sont également imposées pour que le rôle de marquage dévolu à la voie holistique dont font partie ces agents, soit correctement rempli. Afin d'obtenir un traitement suffisamment rapide, la détection des noyaux vocaliques se fait uniquement par une recherche de "patterns" (pics et plateaux) sur la courbe d'énergie, paramètre qui est en outre le premier disponible à la sortie des agents de bas niveau dans le Méta-KS "indices".

Cet agent a été évalué à l'aide de deux tests de robustesse : le premier a été effectué sur un corpus lu extrait de BDFON [Sérignat 92], le deuxième a été effectué sur un corpus de dialogue enregistré dans un environnement bruité. Les résultats sont satisfaisants : (a) 1,6% de noyaux ne sont pas détectés et 4,2% de noyaux sont excédentaires.

4.2. KS pseudo-syllabes

Cet agent rassemble les données sur les noyaux vocaliques et s'assure que la durée entre deux noyaux peut contenir une syllabe en fonction du débit qui est estimé à chaque instant. Si la durée est trop courte il élimine un noyau vocalique et envoie un message au KP noyaux vocaliques pour lui donner cette information et pour lui permettre en retour de réajuster certains de ses seuils de décision.

4.3. KP démarcation de mots

Cet agent est chargé de placer des marqueurs de début et de fin de mot par rapport aux noyaux vocaliques ; s'il le peut, il doit aussi qualifier les mots marqués (en termes de mot lexical ou de mot grammatical) pour aider l'analyse lexicale. Cette tâche de démarcation est réalisée par l'application d'un ensemble de règles prosodiques issues des travaux réalisés dans ce domaine par [Nasri 89] à la suite de l'expertise fournie par [Caelen-Haumont 93]. Le taux

de détection est d'environ 40% avec une fiabilité de 95%. Les résultats sont écrits dans le Méta-KS "limites de mots" qui les propage dans la voie gauche et vers le KP "démarcation de groupes syntaxiques".

4.3. KP démarcation de groupes syntaxiques

Cet agent doit déterminer les début et fin de groupes syntaxiques à partir des mots démarqués par le KP démarcation de mots et des informations fournies par le KP noyaux vocaliques. Il doit aussi déterminer le type des groupes repérés : groupe majeur ou mineur dans la phrase. Il utilise des règles similaires à celle du KP démarcation de mots et qui ont fait l'objet de la même expertise. Il vérifie les frontières de mots donnés par le KS démarcation, propage des contraintes venant des actes de dialogue et fournit des informations aux analyseurs linguistiques qu'il écrit dans le Méta-KS "limites de mots".

Les contraintes fixés à ces deux agents leur imposent de fournir des données fiables pour l'ancrage. Pour cela un seuil de crédibilité est fixé à une valeur suffisante de manière à ce que les résultats retenus ne souffrent d'aucune ambiguïté. Certes, si l'on se prive ainsi de nombreux points d'ancrage potentiels, on n'amorce pas de chemin de recherche incertain et on n'induit pas les autres agents en erreur. Par cette affirmation nous anticipons par là sur la réponse que nous nous sommes promis de donner à propos du problème de la régulation : ce que nous permet de mettre en évidence le module prosodique est bien de savoir si les modules doivent travailler sur le mode de la compétition ou sur celui de la coopération. Nous approfondissons cette question maintenant.

5. La question de la régulation dans le système MICRO

Dans une architecture multi-agents, la maîtrise des processus de régulation¹ revêt une importance primordiale : en effet il ne peut y avoir d'émergence naturelle dans un système tel que MICRO — et probablement dans tout système de compréhension de la parole — du fait de la non-hiérarchie des niveaux et de leur hétérogénéité qui ne donne pas à aucun module le droit de trancher en cas de conflit. Les deux seules solutions sont donc :

(a) la compétition, qui signifie de donner la décision au module qui fournit une réponse en premier, ce qui dans notre cas donnerait systématiquement ce privilège à la voie droite. Il est évident qu'il n'y a aucune raison de procéder ainsi, ni même de privilégier quelqu'agent que ce soit puisqu'aucun n'est infaillible,

(b) la coopération, qui signifie qu'un consensus doit être trouvé entre tous les agents pour obtenir un résultat : c'est le processus d'unification que nous avons évoqué par anticipation plus haut.

Les neurosciences et la psychologie cognitive n'apportant que peu d'indications sur un tel processus de régulation chez l'humain, nous avons tenté de définir des stratégies de contrôle par une simulation par un expert humain en faisant fonctionner le système en pas-à-pas .

Dans une architecture multi-agents, la maîtrise des processus de contrôle est de la première importance. Ce contrôle, bien que localement distribué, apparaît nécessaire à la fois pour gouverner les stratégies d'adaptation contextuelle des agents et pour réguler les stratégies locales d'unification des hypothèses. Venant de rejeter le principe de compétition entre agents, nous en sommes réduits à utiliser la seule hypothèse disponible, la coopération. En reprenant

¹ par contrôle nous entendons surtout contrôle de la propagation des hypothèses et des mécanismes de filtrage des erreurs d'un agent au suivant. En effet l'ensemble du système fonctionnant sous l'angle de la coopérativité entre agents, ceux-ci doivent être garantis de traiter une information suffisamment crédible, s'ils veulent raisonner efficacement.

les principes de Grice [Grice 75] à notre propre compte dans ce cadre d'usage de la coopérativité nous pouvons poser les principes suivants comme maximes de coopération :

- ne pas induire un agent en erreur (maxime de qualité),
- lui fournir une liste d'hypothèses suffisamment ouverte mais pas trop longue (compromis fiabilité/économie du maxime de quantité),
- gérer les conflits en commun (consensus ou vote).

Les agents doivent donc fournir des informations les plus certaines possibles pour ne pas engendrer de fauses décisions ou de faux calculs chez les agents qui reçoivent leurs messages. Cela implique qu'ils doivent faire un contrôle de cohérence et de fiabilité avant de transmettre leurs hypothèses, ce qui complique singulièrement leur tâche et leur pose peut-être un problème de compétence. Pour continuer à explorer cette voie nous avons regardé ces points à la lumière d'une série d'expertises humaines puisque la réalisation des agents se complique singulièrement et que nous devons améliorer nos propres connaissances sur les capacités d'adaptation des traitements des agents ainsi que sur les stratégies d'unification des hypothèses.

5.1. Description de la simulation et résultats locaux

Deux rôles ont été dévolus à l'expert : (a) simuler le fonctionnement des agents et (b) se conformer à une stratégie d'unification coopérative des hypothèses diffusantes. L'expert a travaillé sous différentes contraintes. Rien n'assure cependant que la meilleure stratégie pour un système automatique corresponde à celle employée par l'expert humain. Mais cela donne une première idée avant l'implémentation. L'expertise a porté sur un corpus extrait de BDSONS [Serignat 92], la tâche de l'expert était de faire fonctionner le système pas-à-pas et de propager ou de bloquer des informations.

Nous nous intéressons ici au module "acoustico-phonétique" pour illustrer l'expertise. A ce niveau, nous avons comptabilisé le nombre de fois où l'expert a certifié un trait :

Nom du trait ¹	Utilisation
Macro-classe	100 %
Trait 1	75,4 %
Trait 2	36,8 %
Trait 3	3,5 %
Trait 4	0,0 %
Voisement	96,5 %
Energie	86,0 %

A partir de ce tableau peut tirer les conclusions suivantes sur le principe de quantité de Grice :

- Deux traits principaux semblent fonder la prise de décision, les suivants peuvent ajouter du bruit s'ils ne sont pas assez robustes,
- Les traits "Macro-classe", "Voisement" et "Energie" sont des paramètres robustes. Pour l'expert, les traits "Voisement" et "Energie" n'apportent souvent qu'une confirmation (forte redondance) à l'information portée par "Macro-classe". Cela peut suggérer l'implémentation d'un agent de vérification de traits phonétiques.

Le taux de confiance estimé de l'expert a été :

¹ le trait "Durée intrinsèque" n'a pas été analysé, la trace ne fournissant pas une connaissance suffisante sur son utilisation. Un enregistrement audio de l'expert est nécessaire pour permettre une telle analyse.

- *Confiance forte (C++)*: lors de l'unification, l'expert a conservé uniquement les hypothèses produites par le KP phonologie qui étaient compatibles avec le trait considéré,
- *Confiance ordinaire (C+)*: en sus des hypothèses précédentes, l'expert a conservé celles qui ne contredisaient pas ce trait,
- *Confiance faible (C-)*: l'expert a conservé au moins une hypothèse entrant en contradiction avec le trait considéré.

Enfin, on calcule un taux de confiance "Cm" cumulant les taux de confiances forte et ordinaire. Le tableau suivant présente, en pourcentages, les résultats d'analyse obtenus :

Nom du trait	C++	C+	Cm	C-
Macro-classe	78,9	3,5	82,4	17,6(1)
Trait 1	26,8	22,0	48,8	51,2
Trait 2	23,8	23,8	47,6	52,4
Voisement	92,7	-	92,7	7,3
Energie	83,7	-	83,7	16,3 (2)

(1) rencontré en cas de conflit avec le KP Phonologie, ou avec le trait "Energie" (2) conflit avec "Macro-Classe"

On remarque que:

- l'expert accorde une grande confiance aux traits ("Macro-classe", "Energie" et "Voisement"), à l'exception des cas où ces derniers se contredisent mutuellement. Si de plus, aucun candidat du KP Phonologie n'est compatible avec ces traits, toutes les hypothèses sont conservées. Dorénavant ces traits sont appelés *traits principaux*.
- l'expert accorde une confiance limitée aux autres traits sauf peut-être dans des situations prototypiques simples. Dorénavant ces traits sont appelés *traits secondaires*.

5.2. Stratégies d'unification de l'expert

L'analyse ci-dessus nous donne une idée assez fidèle de la stratégie d'unification employée par l'expert : généralement, celui-ci préfère adopter une stratégie coopérative mais prudente qui consiste à garder un maximum d'hypothèses, en éliminant progressivement les hypothèses les plus mauvaises ou redondantes. La stratégie générale d'unification correspond à un filtrage des hypothèses par les traits "Macro-classe", "Voisement" et "Energie". Dans de rares cas contextuellement caractérisés, l'expert a choisi une stratégie différente. Ce sont ces cas qui nous intéressent maintenant.

Nous avons demandé à l'expert de reprendre ses expertises d'unification en lui imposant tour à tour diverses stratégies systématisées afin de les comparer. Trois stratégies ont ainsi été définies : en premier lieu, une stratégie normale, proche de celle employée précédemment. En second lieu, deux stratégies dites stricte et large qui sont respectivement plus et moins contraignantes que la première stratégie (en fait la stratégie "stricte" respecte davantage la maxime de quantité de Grice). Les trois stratégies se distinguent dans le contexte d'analyse le plus fréquent, où l'on rencontre une forte compatibilité entre les traits. On définit alors les trois stratégies comme suit:

- *stratégie normale* : l'expert garde une hypothèse compatible avec tous les traits principaux et deux traits secondaires,
- *stratégie large* : l'expert garde une hypothèse compatible avec tous les traits principaux,
- *stratégie stricte* : l'expert garde une hypothèse compatible avec tous les traits.

Les autres contextes d'analyse correspondent à des cas particuliers. Les stratégies sont alors identiques. En premier lieu, s'il y a incompatibilité entre les traits eux-mêmes, on détend les contraintes sur les traits secondaires jusqu'à atteindre un point de compatibilité faible. Enfin, si aucune hypothèse phonémique n'est compatible avec les traits, on conserve alors toutes les hypothèses. Ainsi, la confiance ultime est-elle accordée en tout dernier recours aux hypothèses descendantes. Malheureusement, il se peut que celles-ci peuvent avoir été auparavant orientées dans une mauvaise direction par une production ascendante défectueuse.

5.3. Analyse qualitative des stratégies

Nous avons caractérisé la qualité des trois stratégies à l'aide de deux critères : la robustesse des hypothèses produites et l'explosion combinatoire générée.

- *Robustesse* : nous avons observé la robustesse des stratégies au regard du module. Celle-ci est caractérisée par la qualité des propositions au sortir du module, c'est-à-dire les hypothèses lexicales filtrées à partir des unités phonétiques d'unification¹. Le tableau suivant présente les taux de robustesse correspondant à chaque stratégie : celui-ci est bien entendu inversement proportionnel au caractère contraignant de chaque stratégie,

Stratégie	Taux de robustesse
Stricte	62,5 %
Normale	75,0 %
Large	87,5 %

- *Explosion combinatoire* : ce second critère a été évalué à partir du nombre d'hypothèses lexicales conservées à chaque fin d'analyse. Cet indice est d'importance puisqu'un grand nombre d'hypothèses nécessite des efforts de désambiguïsation soutenus de la part du module linguistique. Il en résultera un accroissement des risques d'erreurs et de temps de calcul. Les résultats sont les suivants :

Stratégie	Production multi-hypothèses	Nb hypothèses finales (1)
Stricte	12,5 %	1,2
Normale	37,5 %	1,4
Large	62,5 %	1,9

(1) En moyenne. Le facteur de branchement à chacun des 3 premiers phonèmes des unités lexicales du corpus est de 5.

Là encore, on retrouve des résultats attendus : une stratégie large est source d'une plus grande explosion combinatoire, c'est bien celle qui s'éloigne le plus de la maxime de quantité, mais c'est la plus robuste. On voit donc déjà que les principes de Grice ne peuvent pas être appliqués strictement lorsque les connaissances ne sont pas certaines d'une part et qu'il n'existe pas de stratégie coopérative optimale dans notre cas.

5.3. Guides pour la conception du système

¹ Dans le cas de la simulation, la qualité des hypothèses lexicales reflète totalement celle du groupe acoustico-phonétique, car le groupe linguistique ne fournit ici aucune contrainte sur celles-ci.

La stratégie coopérative semble convenir mais les modalités d'application doivent être adaptées à la situation. En examinant de près les cas où ces performances sont les meilleures, nous avons ainsi déterminé quatre ensembles de paramètres pour l'adaptation contextuelle de la stratégie. Ce sont :

- les paramètres d'environnement acoustique,
- la concordance informative entre les traits,
- le type de phonotype candidat,
- la trace du comportement des agents KP.

En jouant sur ces paramètres il est possible d'obtenir des critères de commande de la stratégie d'unification et des règles de régulation de ce module. Ces règles conditionnent le processus d'unification.

Concordance des traits

La concordance des traits est un effet renforçateur qui peut être utilisé pour contraindre plus strictement la stratégie d'unification (et à l'opposé la non-concordance pour l'élargir). Ainsi l'analyse statistique des expertises montre que l'accord des traits "Macro-classe" et "Energie" donne une robustesse de 96,9%.

Type du phonotype candidat

La robustesse des méthodes de décodage acoustico-phonétique varie fortement avec le type d'unité phonétique attendu. Typiquement, les fricatives sont très bien identifiées tandis que les consonnes liquides posent toujours problème. La règle d'adaptation est donc :

Si X est un phonotype candidat et si X présente des traits de:

- *Fricative*: alors stratégie_stricte (le trait strident est très robuste dans ce cas).
- *Occlusive ou Voyelle*: alors stratégie_normale (confiance moyenne dans les traits).
- *Autre* alors stratégie_large

6. Conclusion

L'architecture distribuée de *MICRO* permet une modélisation modulaire ainsi que la co-existence de deux types d'analyses parallèles complémentaires (analytique et holistique). Ces caractéristiques permettent d'espérer une amélioration de la robustesse de l'analyse, mais posent de nombreux problèmes de régulation (pour ne pas dire de contrôle). C'est la raison pour laquelle nous avons procédé à une simulation d'une partie du système. La principale conclusion de cette simulation réside dans la nécessité d'adopter une stratégie coopérative adaptative, orientée par la crédibilité des hypothèses transmises par un agent. Cela pose la question de l'organisation "sociale" du travail des agents et du raisonnement collectif à partir d'hypothèses incertaines. Sur un plan de génie logiciel ces architectures semblent intéressantes : structure unifiante de modules hétérogènes, répartition des charges, distribution des fonctions, meilleure sécurité de fonctionnement, performances accrues grâce au parallélisme. Mais de nombreux progrès restent encore à faire pour réaliser des systèmes opérationnels. Cette étude permet cependant d'entrevoir des solutions originales pour un système de compréhension de la parole entièrement intégré.

Remerciements

Le travail présenté ci-dessus est un travail d'équipe auquel ont participé J.Y. Antoine, O. Baujard, G. Caelen-Haumont, B. Caillaud, O. Delemar, C. Garbay, H. Kabré et P. Munteanu. Qu'ils en soient chaleureusement remerciés. Ce travail a reçu également le soutien du GDR-PRC Communication Homme-Machine.

7. Bibliographie

Références sur les approches cognitives

- J.Y. Antoine, B. Caillaud, J. Caelen (1993), "Syntax-semantics cooperation in MICRO, a multi-agent speech understanding system", *EUROSPEECH'93, Berlin*, 1993, pp. 2139-2142.
- G.T.M. Alman (1990), "Introduction" in G. Altman (Ed.) *Cognitive models of speech processing*, MIT Press, Cambridge, Mass.
- O. Baujard, C. Garbay (1990), "A programming environment for distributed expert system design", *Expert System Applications, ExperSys*, pp. 27-32.
- J. Caelen (1979), *Un modèle d'oreille. Analyse de la parole. Reconnaissance phonémique*, Thèse d'Etat, Toulouse, Vol. I & II.
- J. Caelen et al. (1990), "Architecture et fonctionnement du système DIRA. De l'acoustique aux niveaux linguistiques", *Revue de traitement du signal*, vol. 7, no 4, pp. 345-366.
- J. Caelen, C. Garbay (1992), "Modularité pour la reconnaissance de la parole", *ECCOS'92*, Orsay, 1992, pp. 44-63.
- G. Caelen-Haumont (1993), "Cognitive processes and prosodic encoding. Speaker's adaptation to discourse conditions", *Communication and cognition - Artificial Intelligence*, 10:4.
- J. Fodor (1983), *The Modularity of Mind*, MIT Press, Cambridge MA.
- R.W. Gardner (1981), *The analysis of unanalysed perceptions*, in M. Kubovy, J.R. Pomerantz (eds.), "Perception and organisation", Lawrence Erlbaum, Hillsdale NJ, pp. 119-139.
- S.M. Kosslyn (1980), *Image and Mind*, Harvard University Press, Cambridge MA.
- J. Lautrey (1990), *Esquisse d'un modèle pluraliste du développement cognitif*, in Reuchlin et al (eds), "Cognition, l'individuel et l'universel", PUF, Paris.
- J.L. McClelland et J.L. Elman (1986), "The TRACE model of speech perception", *Cognitive Psychology*, 18, pp. 1-86.
- D. Marr (1982), *Vision: a Computational investigation into the human representation and processing of visual information*, Freeman, San Francisco.
- W.D. Marslen-Wilson (1987), "Functional parallelism in spoken word recognition", *Science*, 25, pp. 71-102.
- J. Mehler et E. Dupoux (1990), *Naître humain*, Odile Jacob, Paris.
- M. Minsky (1985), *The society of Mind*, Simon & Schuster, New-York.
- C. Morris (1946), *Signs, language and behavior*, Prentice Hall, New-York.
- J. Morton (1984), *Brain-based and non-brain-based models of language*, in D. Caplan, L. Lecours, A. Smith (eds), "Biological Perspectives in Language", MIT Press, Cambridge MA, pp. 40-64.
- P. Munteanu (1992), "Modification d'un gestionnaire de connaissances sur la parole", Rapport de stage ingénieur de l'Institut polytechnique de Bucarest-Roumanie, ICP/INPG, Grenoble.
- M.K. Nasri, G. Caelen-Haumont et J. Caelen (1989), "Using Prosodic Rules in Continuous Speech Recognition", *Proc. of ICASSP-IEEE*, Glasgow, Vol. 1, 1989, pp. 671-674.
- F. Rastier (1987), "Sémantique et Intelligence Artificielle: Présentation", *Langages*, 87, pp. 5-19.

J.F. Sérygnat, J.M. Dolmazon, M. De Calmes, G. Pérennou (1992), "BDSONS - BDLEX : Bases de données vocales et lexicales", *Rencontres Recherches en Informatique - Industrie*, Toulouse, 1992, pp. 88-89.

R.W. Sperry (1961), "Cerebral Organization and Behavior", *Science*, 133, pp. 1749-1757.

Articles fondateurs sur les systèmes de reconnaissance

J. Baker (1975), "The DRAGON system - an overview". *IEEE Trans. ASSP-23*, 1, pp. 24-29.

J. Caelen, M.K. Nasri, E. Reynier, H. Tattegrain (1990), "Architecture et fonctionnement du système DIRA. De l'acoustique aux niveaux linguistiques". *Revue TS*, vol. 7 n° 4, pp. 345-366.

N. Carbonell, J.P. Damestoy, D. Fohr, J.P. Haton, F. Lonchamp (1986a), "Design and implementation of an acoustic-phonetic decoding expert system". *IEEE-ICASSP*, Tokyo, Japan, 1986, pp. 392-397.

N. Carbonell, J.M. Pierrel (1986b), "Architecture and knowledge sources in human computer oral dialogue system". *Proceedings of NATO Workshop*, Corsica, France, 1986.

Y. Chow et al (1987), "BYBLOS: The BBN continuous speech recognition system". *IEEE ICASSP*, 1987, pp. 238-242.

H.G. Goldberg, R. Reddy (1976), "Feature extraction, segmentation and labeling in the Harpy and Hearsay-II systems". *J. Acoust. Soc. Am.* Vol. 60, S11 (A), pp. 1044-1067.

J.P. Haton (1985), "Intelligence artificielle en compréhension automatique de la parole: état des recherches et comparaison avec la vision par ordinateur". *TSI*, Vol. 4-3, pp. 265-287.

D.H. Klatt (1975), *Word verification in speech understanding system* in "Trends in speech recognition" Reddy ed., Academic Press, New York, pp. 321-341

D.H. Klatt (1977), "Review of the ARPA speech understanding project". *J.A.S.A.*, Vol. 62, No. 6, pp. 2405-2420.

F. Kubala et al (1988), "Continuous speech recognition results of the BYBLOS system on the DARPA 1000-word resource management database". *IEEE ICASSP*, pp. 291-294.

K.F. Lee (1988), *Large vocabulary speaker-independent continuous speech recognition : The SPHINX system*. Ph.D. dissertation, Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

V.R. Lesser, R.D. Fennel, L.D. Erman, R. Reddy (1975), "Organization of the Hearsay-II speech understanding system". *IEEE Trans. Acoust. Speech Signal Process.* ASSP-23, pp.11-23.

V.R. Lesser, L.D. Erman (1977), "A retrospective view of the Hearsay-II architecture". *IJCAI-77*, 1977, pp. 210-216.

B.T. Lowerre (1976), *The Harpy speech recognition system*. Ph.D. thesis (departement of computer science, Carnegie-Melon University), 1976.

J. Mariani, J.S. Lienard (1978), "ESOPE 0 : Un programme de compréhension automatique de la parole procédant par prédiction-vérification". *Actes du Congrès AFCET-IRIA RFIA.*, 1978, pp. 123-129.

J. Mariani (1982), "The ESOPE continuous speech understanding system". *IEEE-ICASSP*, pp. 1637-1640.

G. Mercier, P. Quinton, R. Vives (1977), "Dialogue homme-machine avec KEAL". *Recherches Acoustiques*, Vol. 4, pp. 187-206.

G. Mercier et al (1980), *The KEAL speech understanding system*. In *Spoken Language Generation and Understanding*, J.C. Simon, editor, D. Reidel publishing.

A. Newell, J. Barnett, J.W. Forgie, C.C. Green, D.H. Klatt, J.C. Licklider, J. Munson, R. Reddy, W.A. Woods (1973), *Speech understanding systems*. NORTH-HOLLAND Press, Amsterdam.

G. Perennou (1980). "ARIAL II: System for speech recognition". *5th IJCP*, Miami, December 1980, pp. 83-91.

J.M. Pierrel (1981), *Etude et mise en oeuvre de contraintes linguistiques en compréhension automatique du discours continu*. Thèse d'Etat, Université de Nancy 1.

- J.M. Pierrel (1982), "Utilisation des contraintes linguistique en compréhension automatique de la parole continue". *TSI*, vol. 1, n°5, pp. 403-421.
- R. Reedy, L.D. Erman, R.B. Neely (1973), "A model and a system for machine recognition of speech". *IEEE Trans. AU-21*, pp. 229-238.
- R. Reddy, L.D. Erman (1975), *Tutorial on system organization for speech understanding*. Academic Press, New York, pp. 457-459.
- B. Ritea (1975), "Automatic speech understanding systems". *Proceedings of the 11th IEEE Computer society Conference*, Washington, DC, 1975, pp. 319-322.
- R.Y. Sheryl, G.H. Alexander, H.W. Wayne, T.S. Edward, W Philippe (1989), "High level knowledge sources in usable speech recognition systems". *Communications of the ACM*, Vol 32, No. 2, pp 78-90.
- W.A. Woods (1970), "Transition network grammars for natural language analysis". *Commun. Assoc. Comput. 13*, pp. 561-602.
- W. Woods, M. Bates, G. Brown, B. Bruce, C. Cook, J. Klovstad, J., Makoul, B. Nash-Webber, R. Schwartz, J. Wolf, V. Zue, V. (1976), *Speech understanding systems : final technical progress report*. Bolt Beranek and Newman, Inc. Cambridge.