ORSAY

N° D'ORDRE : 5804

UNIVERSITÉ DE PARIS-SUD
U.F.R. SCIENTIFIQUE D'ORSAY

THÈSE

présentée

Pour obtenir

Le GRADE de DOCTEUR EN SCIENCES
DE L'UNIVERSITÉ PARIS XI ORSAY

PAR

Andrei POPESCU-BELIS

SUJET : **Modélisation multi-agent des échanges langagiers : application au problème de la référence et à son évaluation**

Soutenue le 2 juillet 1999 devant la Commission d'examen

M.	Joseph MARIANI	Examinateur
M.	Jean-Marie PIERREL	Rapporteur
Mme	Anne REBOUL	Examinateur
M.	Gérard SABAH	Directeur
M.	Luc STEELS	Rapporteur
M.	Bernard VICTORRI	Examinateur

REMERCIEMENTS

Je voudrais tout d'abord remercier vivement Gérard Sabah, qui a constamment veillé à ce que cette thèse se déroule dans les meilleures conditions, et lui dire tout le plaisir que j'ai eu à travailler sur les thèmes que j'ai pu découvrir et approfondir grâce à lui.

Mes plus sincères remerciements s'adressent ensuite aux membres du jury de thèse, en commençant par Joseph Mariani, président du jury, dont les actions en faveur de l'évaluation ont grandement influencé mon travail.

Je voudrais exprimer toute ma gratitude à Jean-Marie Pierrel pour son chaleureux soutien à plusieurs moments clé, et pour avoir accepté d'être rapporteur de cette thèse.

Ma reconnaissance va également à Anne Reboul, pour ses nombreux conseils et les discussions cordiales autour du projet CERVICAL, dont l'influence sur une grande partie de cette thèse a été déterminante.

Les travaux de Luc Steels ont toujours représenté pour moi une référence et un modèle ; je le remercie sincèrement pour l'honneur qu'il m'a fait d'être rapporteur de cette thèse.

Je voudrais également remercier chaleureusement Bernard Victorri, dont les exposés, et les échanges que nous avons eus, ont contribué à l'ouverture de cette thèse vers certains débats actuels en sciences cognitives.

Cette thèse doit beaucoup à l'environnement remarquable du LIMSI-CNRS : je voudrais à ce titre exprimer ma profonde gratitude à l'ensemble de son personnel scientifique, technique et administratif, qui contribue à faire de ce laboratoire un endroit extrêmement propice à la recherche.

Je remercie très sincèrement Isabelle Robba pour notre collaboration, qui a représenté pour moi l'une des étapes les plus enrichissantes et les plus agréables de mon travail. Je remercie également mon ami Franck Ramus pour nos nombreuses et instructives discussions scientifiques.

La plupart des membres du groupe Langage et Cognition ont influencé, d'une façon plus ou moins directe, l'élaboration de cette thèse, et je leur suis pleinement reconnaissant. Parmi eux, je citerai plus particulièrement Xavier Briffault, Olivier Ferret, Benoît Habert et Christian Jacquemin, ainsi que Patrick Paroubek (groupe TLP).

Enfin, je voudrais exprimer ma reconnaissance à tous ceux qui ont été proches de moi pendant ces années de thèse, en dépit de mon manque de temps chronique : je pense à mes parents, à mes amis, et à Cécile.

SOMMAIRE

Introduction *5*

Première partie : vers une société d'agents utilisant le langage **9**

Chapitre I *Approche symbolique et nouvelles voies en intelligence artificielle et traitement automatique des langues* *11*

Chapitre II *Réalités et simulations de la communication langagière* *25*

Chapitre III *Principes de modélisation : la « pragmatique simulée »* *69*

Chapitre IV *Deux réalisations : études d'un agent et d'une société* *83*

Deuxième partie : entités, représentations, référence **115**

Chapitre V *La référence : théories, modèles, systèmes* *116*

Chapitre VI *Du fonctionnement de la référence à la capacité référentielle d'un agent* *161*

Chapitre VII *Résolution de la référence : un modèle représentationnel et son implémentation* *197*

Troisième partie : l'évaluation en traitement automatique des langues **241**

Chapitre VIII *Cadre théorique et conditions de cohérence des mesures d'évaluation* *242*

Chapitre IX *Critiques et propositions pour évaluer le traitement de la référence* *257*

Chapitre X *Conclusion d'ensemble* *293*

Chapitre XI *Annexes* *297*

Références bibliographiques *315*

Table des matières détaillée *333*

INTRODUCTION

Les possibilités des ordinateurs à traiter les langues naturelles ont été explorées dès les premières réalisations informatiques. Manifestation par excellence de « l'intelligence », la faculté linguistique était conçue aussi comme un moyen de tester « l'intelligence artificielle » sans faire appel à des interfaces autres qu'un simple terminal texte. Le jeu de l'imitation proposé par Turing (1950) consiste, dans cet esprit, à dialoguer sur un terminal avec un ordinateur et un humain, et à essayer de différencier les deux. L'optimisme initial est visible dans cette prédiction de Turing :

> *Je crois que dans une cinquantaine d'années on pourra programmer des ordinateurs [...] de façon à ce qu'au jeu de l'imitation un interrogateur moyen n'aura pas plus de 70% de chances de repérer l'ordinateur au bout de cinq minutes.*

À la confiance initiale dans les potentialités des machines à utiliser le langage ont succédé de vives critiques, puis un certain oubli du test de Turing et l'étude de tâches linguistiques plus applicatives. Le bilan actuel doit être pour le moins nuancé. Alors que Turing donnait à l'arbitre de son test moins de 70% de chances d'identifier la machine (50% correspondent au hasard, donc à l'imitation parfaite), les performances actuelles se situent entre 85 et 90%[1]. Sur de nombreuses autres tâches, les programmes de *traitement automatique des langues* (TAL) atteignent le seuil d'utilité, voire de rentabilité.

Il reste que la plupart des réalisations en TAL reposent sur la conception traditionnelle du langage comme système de symboles, la décomposition de son traitement en couches (lexique/syntaxe/sémantique/…), et leur formalisation computationnelle de type logique. En un certain sens, l'idée – présente chez Turing – de doter la machine de compétences linguistiques seulement à travers un terminal texte domine encore les recherches en TAL.

C'est une dimension supplémentaire que notre thèse cherche à apporter au paradigme symbolique actuel. Nous cherchons en effet à reproduire le caractère *situé* du langage, à savoir dépendant d'un *environnement*, d'un support biologique *individuel*, et d'une fonction et évolution *collectives*. À ce changement de perspective correspondent nos propositions concrètes de modélisation, provenant d'un examen raisonné des bases naturelles du langage et de son fonctionnement. En particulier, nous défendons de façon

[1] Donc la machine est bien plus reconnaissable que ne l'espérait Turing. Le *Prix Loebner* est en effet décerné chaque année depuis 1990 au programme obtenant le meilleur score lors d'une session officielle du « test de Turing » (http://www.loebner.net/Prizef/loebner-prize.html). Il faut signaler que les cinq minutes disponibles sont un facteur très limitant.

originale l'importance du phénomène de la *référence* dans cette modélisation, à savoir le lien réalisé par le langage entre les *entités du monde de référence* et les *représentations des individus ou agents*.

L'argumentation et la description des réalisations s'organisent en trois parties. La première partie étudie dans sa généralité le paradigme de la « pragmatique simulée » que nous proposons.

- Le premier chapitre présente différentes tentatives de dépassement du paradigme symbolique en TAL, montrant ainsi les points sur lesquels celui-ci peut être amélioré.
- Le deuxième chapitre étudie les sources d'inspiration naturelles pour deux directions clé de notre approche : l'ancrage biologique du langage dans l'individu, et son ancrage collectif, en particulier trois cas d'émergence. Des modèles informatiques de ces phénomènes sont étudiés, notamment ceux inspirés par la théorie de Gerald Edelman, ainsi que les différentes modélisations des échanges langagiers à l'aide d'agents.
- Les principes tirés du deuxième chapitre nous permettent d'esquisser, au Chapitre III, un modèle multi-agent en accord avec la « pragmatique simulée ».
- Deux modèles informatiques réalisés comme éléments de base de notre modélisation par agents font l'objet du Chapitre IV. Ils se situent l'un au niveau individuel, l'autre au niveau collectif.

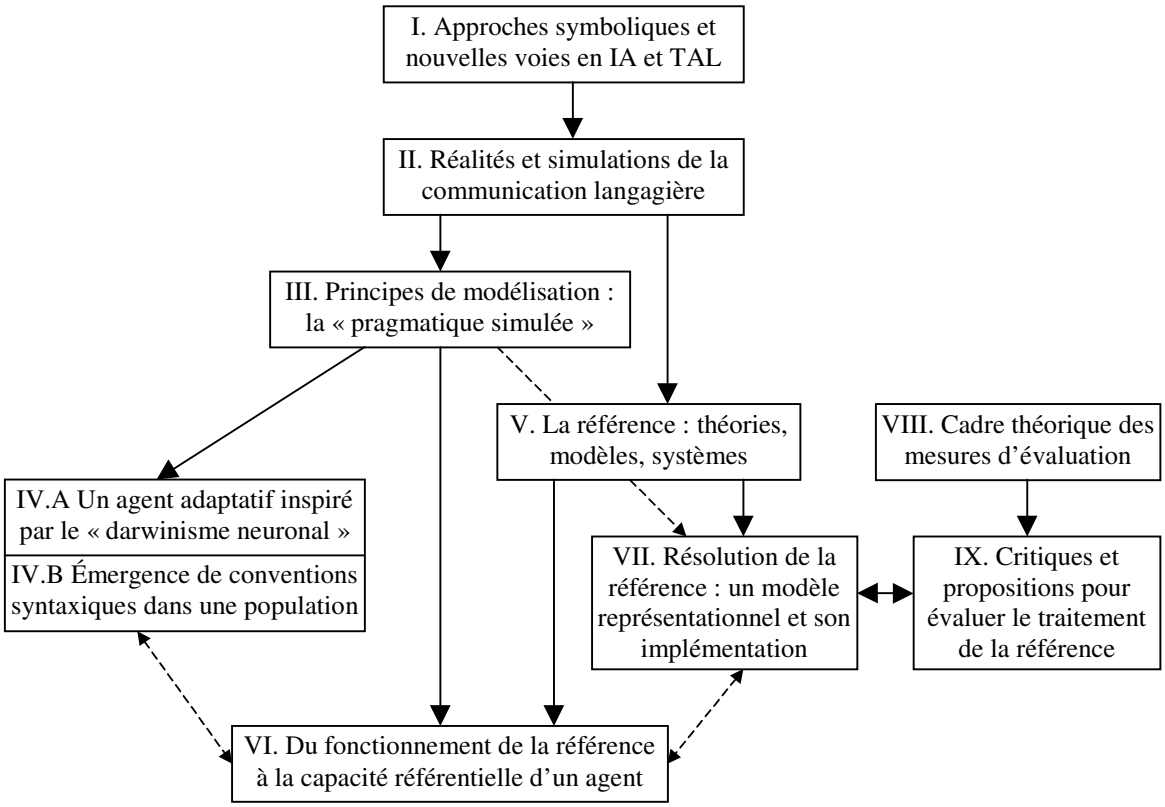
La deuxième partie aborde la question de l'ancrage du langage dans l'environnement par le biais de la *référence linguistique*, c'est-à-dire le lien de désignation, par le biais d'expressions du langage, entre les représentations des locuteurs et les entités du monde de référence.

- Nous commençons par une étude de la nature de la référence, des modèles computationnels réalisés, et des méthodes proposées pour la traiter automatiquement (Chapitre V).
- Ces considérations sur la référence, auxquels s'ajoutent des éléments inspirés de la biologie et de la linguistique fonctionnelle, nous permettent de proposer au Chapitre VI un modèle de la capacité référentielle d'un agent, qui se détache des théories logicistes de la référence.
- Afin d'explorer plus avant l'étude de la référence sur des textes réels, nous proposons un cadre théorique représentationniste compatible avec la pragmatique de la référence, et nous décrivons son implémentation en tant que résolveur de références, ainsi que les outils réalisés et les résultats obtenus (Chapitre VII).

La troisième partie aborde le problème de l'évaluation des programmes de TAL. La « compréhension » ne semble pas pouvoir être évaluée en soi, mais seulement à travers des tâches bien déterminées, qui la sous-entendent plus ou moins directement.

- Nous proposons d'abord, au Chapitre VIII, un modèle générique d'évaluation des capacités d'un programme, modèle permettant de choisir des mesures d'évaluation appropriées.
- Puis, nous appliquons ce modèle à l'analyse de différentes mesures d'évaluation de la référence, soit existantes, soit originales ; nous proposons entre autres une mesure fondée sur le concept d'*information référentielle* (Chapitre IX). Une perspective sur l'évaluation de la modélisation multi-agent vient clore cette partie, suivie d'une conclusion d'ensemble.

Pour une meilleure intelligence de l'ensemble, nous représentons l'articulation des chapitres à l'aide du schéma suivant. Les principaux liens figurent en trait plein.



Cette thèse peut être parcourue d'au moins deux manières. La première consiste à suivre l'ordre naturel, et comprendre ainsi les enjeux et la position défendue ici dans le débat de principe sur les nouvelles voies de l'intelligence artificielle et du traitement automatique des langues. Les deux premières réalisations, puis l'application à la référence et son évaluation, viendront préciser, par touches successives, notre position et le paradigme de la « pragmatique simulée ». Mais on peut aussi, si l'on s'intéresse aux

questions de détail, tirer profit de chaque partie individuellement. C'est le cas des différents états de l'art commentés, que nous avons rendus détaillés et actualisés, et bien sûr aussi de nos différentes propositions, qui s'inscrivent aussi dans des débats propres à chaque domaine.

PREMIÈRE PARTIE : VERS UNE SOCIÉTÉ D'AGENTS UTILISANT LE LANGAGE

Chapitre I

APPROCHE SYMBOLIQUE ET NOUVELLES VOIES EN INTELLIGENCE ARTIFICIELLE ET TRAITEMENT AUTOMATIQUE DES LANGUES

I.A Objectifs et méthodes de l'intelligence artificielle

Dans son acception la plus large, le terme « intelligence artificielle » désigne la science ayant pour objectif la reproduction des *fonctions* et/ou *mécanismes* de l'intelligence humaine, grâce aux ordinateurs. Ces études nécessitent souvent l'apport d'autres sciences, et passent par l'élaboration de modèles abstraits, mais présentent toujours la possibilité d'une implémentation sur des ordinateurs, permettant d'exploiter et d'évaluer les modèles[1].

L'étude des mécanismes de l'intelligence – *comment* celle-ci fonctionne – représente le côté théorique de l'intelligence artificielle (IA), visant à comprendre l'être humain. La reproduction des fonctions de l'intelligence – *ce que* celle-ci permet de faire – représente le côté constructif, dont les applications sont très nombreuses. Le but de l'IA est de répondre à ces deux interrogations.

Chez l'homme, les différentes fonctions de l'intelligence sont réalisées par des mécanismes biologiques, fruit de l'évolution naturelle. Une des premières questions que doit examiner l'IA est le caractère nécessaire du lien entre fonction et mécanisme : combien faut-il imiter le mécanisme naturel afin d'en reproduire la fonction ? Il semble que l'analyse théorique de cette question soit moins efficace que son étude à travers les diverses réalisations ; les succès ou échecs respectifs permettent d'estimer la proportion

[1] Le mot *fonction* doit être entendu, tout au long du chapitre, dans le sens quasi biologique de « manifestation » ou « activité exercée » et non pas dans celui de « rôle » ou « mission ».

D'un point de vue philosophique la notion d'*ordinateur* demanderait une définition : nous nous en tiendrons aux exemples courants, fondés sur l'architecture dite de Von Neumann. Les progrès de la biologie conduiront peut-être à la question : un cerveau développé *in vitro* est-il un *ordinateur* ?

La notion d'*intelligence humaine* demanderait aussi à être précisée. Il s'agit des capacités de traitement de l'information captée par les organes de sens, capacités communes à tous les humains, et dont certaines sont partagées avec les animaux supérieurs, en particulier les primates. Ainsi, ne semblent pas relever de l'IA : la locomotion, l'identification des nutriments, la capacité immunitaire, la reproduction du matériau génétique – bien que ces domaines soient sources d'importantes analogies.

dans laquelle on doit imiter un mécanisme[2]. Une deuxième question est le degré d'indépendance des différentes fonctions subsumées dans la notion d'intelligence ; cette question, plus centrale peut-être pour les sciences de l'homme, peut recevoir aussi une réponse expérimentale grâce à l'IA.

Les débuts de l'intelligence artificielle se confondent avec ceux de l'informatique, puisque l'ordinateur a été conçu comme une « machine intelligente ». Afin de mieux comprendre les grandes orientations de l'IA ainsi que ses options actuelles, il est utile de revenir sur quelques travaux marquants qui ont fixé le paradigme de l'IA dite *symbolique* ou classique.

I.A.1 Travaux fondateurs de Turing

Les principes du paradigme symbolique sont déjà visibles dans les travaux fondateurs d'Alan Turing[3], qui opèrent deux réductions essentielles. D'abord, Turing montre qu'une classe importante d'opérations mathématiques (les « fonctions récursivement calculables », ici *FRC*) peuvent être réalisées par un dispositif physique approprié. La définition des *FRC* est donnée dans un langage formel qui peut exprimer toutes les opérations effectuées par les langages de programmation, y compris les langages actuels[4].

La deuxième réduction est plus audacieuse. Turing fait remarquer que le raisonnement mathématique peut être exprimé dans le langage formel précédent ; on peut transformer un théorème en une assertion sur les nombres, donc en une *FRC*[5]. Les raisonnements mathématiques peuvent donc être réalisés ou simulés par un ordinateur. Ces raisonnements étant une manifestation par excellence de l'intelligence, on pourrait conclure que *tout processus* intelligent peut être simulé sur un ordinateur. Turing défend ainsi la possibilité d'une intelligence artificielle et propose un test pour la détecter (Turing 1950). La réduction de Turing est donc une instance précoce et exemplaire de la réduction de l'intelligence à une manipulation de symboles reproductible par une machine.

---

[2] Deux exemples extérieurs à l'IA illustrent bien ce point : la locomotion des artefacts a été résolue grâce aux roues et non aux jambes ; en revanche, les caméras numériques ont une structure relativement proche de l'œil. Rappelons aussi l'exemple des appareils volants, assez récurrent dans l'épistémologie de l'IA.

[3] Notre article (Popescu-Belis 1995) contient des références à ce sujet classique.

[4] La thèse de Church-Turing affirme que tout langage de programmation peut se réduire à ce langage « standard » comprenant les boucles, les branchements, etc. Bien sûr, il ne faut pas confondre les *méthodes de programmation*, par exemple les réseaux de neurones ou les systèmes experts, avec le *langage* qui permet leur implémentation, par exemple *Pascal*, *C*, *Prolog* ou *LISP*. Par ailleurs, le « dispositif physique approprié », ancêtre des ordinateurs, est appelé aujourd'hui « machine de Turing ».

[5] Par exemple grâce à la réduction de Gödel.

I.A.2 L'intelligence artificielle symbolique

De nombreux systèmes qui mettent en œuvre la réduction précédente sur différents problèmes ont vu le jour dans les années 1960-1980. L'hypothèse qui sous-tend le *paradigme symbolique* en IA a été le plus clairement formulée par Alan Newell et Herbert Simon (Newell 1980, Newell et Simon 1976). Il s'agit de *l'hypothèse du système physique de symboles*[6] *(SPS)* :

> *Un SPS consiste en un ensemble d'entités, appelées symboles, qui sont des formes physiques pouvant apparaître en tant que composants d'un autre type d'entités, appelées expressions (ou structures de symboles). Ainsi, une structure de symboles est composée d'un certain nombre d'instances de symboles reliées d'une manière physique. […] Le SPS contient aussi un ensemble de processus qui opèrent sur des expressions pour produire d'autres expressions : création, modification, reproduction et destruction. Un SPS est une machine qui produit un ensemble de structures de symboles évoluant dans le temps. […]*
> *Un SPS possède les moyens nécessaires et suffisants pour l'action intelligente en général.* (Newell et Simon 1976:161, nous traduisons)

Outre cette définition du paradigme symbolique, Newell a formulé d'une manière plus précise la réduction de Turing de la pensée à la machine, en décrivant le principe de la programmation en couches (Newell 1982). On distingue dans l'ordre les couches : atomique, cristalline, électronique, logique, puis celles du langage d'assemblage, du langage de programmation, du langage symbolique et enfin la plus élevée et originale, la *strate de connaissance* (« *knowledge level* »). À ce dernier niveau, on rencontre des agents rationnels, manipulant des connaissances, selon un principe de rationalité : il s'agit, selon Newell, du domaine de l'intelligence artificielle.

Dans une optique plus pragmatique, on définit souvent l'IA comme l'ensemble des techniques qui permettent la résolution des problèmes pour lesquels aucune solution n'est connue, qui soit algorithmique et d'une durée raisonnable (Laurière 1987). Cette définition ne présuppose pas nécessairement le cadre symbolique – sauf à restreindre la notion de « problème » aux problèmes symboliques – mais l'ouvrage cité précise que les informations traitées sont toujours symboliques (Laurière 1987:3). Il semble alors que l'augmentation de la vitesse des ordinateurs puisse faire sortir du champ de l'IA certains problèmes, comme peut-être aujourd'hui le jeu d'échecs ; la frontière entre intelligence artificielle et informatique avancée pourrait ainsi s'estomper.

I.A.3 Domaines d'application et réalisations de l'IA

Historiquement, les pionniers de l'IA ont abordé des questions comme la démonstration automatique de théorèmes[7], le jeu d'échecs, voire la résolution générique

[6] En anglais « physical symbol system hypothesis ».

[7] Où l'on retrouve l'aspiration de Turing.

de problèmes (années soixante). Avec la naissance des systèmes experts dans les années soixante-dix, et les travaux logiques autour de la formalisation des « connaissances », le répertoire de l'IA s'est considérablement élargi, et en même temps les espoirs de la communauté. Les principaux problèmes relevant de l'IA (Rich 1983:3, Winston 1984:6 sq.) sont alors : les jeux de réflexion (échecs, dames, go, etc.), la preuve de théorèmes, la résolution générique de problèmes, l'expertise spécialisée (systèmes expert utilisés en médecine, chimie, ingénierie, etc.), la perception (vision/reconnaissance des formes, audition/reconnaissance de la parole), la compréhension de la langue.

Une évaluation de ces réalisations dépasse largement notre propos. Signalons simplement que la représentation des connaissances se heurte encore à des difficultés formelles (représentation du sens commun, des catégories, du cadre d'application ou « frame problem », etc.), que la preuve de théorèmes n'a toujours pas atteint le niveau des experts, et que les jeux semblent avoir progressé surtout grâce à la vitesse et à la mémoire des ordinateurs (champion du monde d'échecs, mais niveau très moyen au jeu de go). La perception artificielle s'est sensiblement écartée de l'approche symbolique, et les espoirs initiaux concernant la compréhension du langage (années cinquante) se sont vite heurtés à de sérieuses difficultés (cf. I.B).

Une remarque de principe s'impose : les difficultés de l'approche symbolique n'entraînent pas de fait l'impossibilité d'une intelligence artificielle implémentée sur un ordinateur. Les ordinateurs sont capables de simuler numériquement de nombreux processus physiques ; or, en dernière instance, l'intelligence est probablement aussi un processus physique. Simplement, la réduction de Turing, ou l'hypothèse du système physique de symboles de Newell et Simon, peuvent ne pas être une méthode adéquate de *simulation* de l'intelligence. C'est pourquoi des paradigmes d'implémentation nouveaux, non symboliques, ont été proposés, en particulier pour la compréhension du langage naturel.

I.B L'intelligence artificielle et le langage

La manipulation du langage par les ordinateurs a été rapidement considérée comme un objectif central de l'intelligence artificielle[8] (Sabah 1988:23-28) ; la traduction automatique, en particulier, a été une des premières tâches étudiées. Devant la disparité entre les ambitions et les résultats affichés, le domaine a subi vers 1965 le coup d'arrêt que l'on sait, avant de redémarrer et de représenter aujourd'hui, dans la « société de l'information », une préoccupation majeure.

D'un point de vue conceptuel, le langage humain est une *fonction* indissociable de l'intelligence (cf. note 1), et peut-être indispensable à celle-ci. Par conséquent, le but de l'IA appliquée au langage – c'est-à-dire du traitement automatique des langues (TAL) –

[8] Ce titre ne peut manquer d'évoquer l'ouvrage de Gérard Sabah (1988, 1989), qui expose en détail les méthodes de l'intelligence artificielle classique/symbolique appliquées au traitement automatique des langues.

est de reproduire la *capacité de langage* ainsi que son *mécanisme* grâce à un système informatique. Cet objectif générique s'accompagne de deux questions : (1) quelle est précisément la dépendance entre fonction et mécanisme, pour le langage, ou en d'autres mots, dans quelle mesure faut-il imiter le mécanisme pour obtenir la fonction ? (2) Quelle est la dépendance du langage par rapport aux autres fonctions de l'intelligence, ou encore, combien peut-on isoler le TAL par rapport à l'IA ?

L'intelligence artificielle symbolique utilise l'hypothèse du système physique de symboles pour répondre à la première question. Nous avons vu en effet comment Turing, déjà, pensait réduire les processus intelligents mathématiques ou linguistiques à la manipulation de symboles par l'ordinateur[9]. En réalité, la réduction des formes linguistiques aux formes logico-mathématiques est plus ancienne, car elle appartient au positivisme logique du début du XX^{e} siècle (Frege 1892/1971, Russell 1905/1956). L'ensemble de ces idées, auxquelles s'ajoutent les formalisations linguistiques proposées par Noam Chomsky (1957, 1965), a donné naissance à un ensemble de méthodes devenues classiques pour la représentation symbolique de la langue.

I.B.1 Objectifs du TAL

Les objectifs « fonctionnels » du TAL consistent à réaliser des systèmes capables d'effectuer des tâches linguistiques, analogues à celles des humains, mais plus rapidement, ou sur plus de données, ou pour un prix moins élevé. Parmi les tâches les mieux définies on compte la traduction, le résumé, la commande en langue naturelle, la recherche d'information dans une base de textes. Les performances de certains systèmes atteignent parfois le seuil d'utilité – moteurs de recherche sur Internet – voire le seuil de rentabilité – outils d'aide à la traduction, gestion de la documentation d'entreprise, reconnaissance vocale.

Ces tâches « finales » reposent souvent sur des traitements internes communs, qui ont isolément une utilité moindre. On peut citer l'étiquetage morpho-syntaxique, l'analyse syntaxique, la segmentation, la résolution des anaphores pronominales, etc. (ces tâches étant souvent associées aux niveaux de traitement décrits au paragraphe suivant). Ici, les performances varient beaucoup selon la tâche. Par ailleurs, on peut défendre l'idée que toutes ces tâches sont des conséquences d'une seule capacité humaine, celle de *comprendre* des énoncés en langue naturelle (et la transformation duale qu'est la *génération*). De nombreux débats ont lieu à ce niveau, et nous dirons simplement que du point de vue de l'IA symbolique, cette capacité se se définit comme la manipulation de représentations symboliques du sens d'un texte.

Les objectifs liés à *l'explication* des *fonctions* précédentes semblent plus éloignés, en partie parce que la biologie n'a pas encore pu vérifier la réalité de la réduction symbolique, et la psychologie commence seulement à étudier la nature des représentations internes. On distingue donc parfois le point de vue des « industries de la

[9] Il est significatif de noter que le test proposé par Turing (1950) pour détecter le comportement intelligent passe exclusivement par l'usage sensé de la langue écrite.

langue », intéressées par les fonctions, de celui des « sciences cognitives », visant l'explication ou la modélisation de ces fonctions grâce à l'informatique.

I.B.2 Niveaux de traitement linguistique

La décomposition classique en TAL du traitement des données linguistiques distingue les niveaux lexical, syntaxique, sémantique, pragmatique, discursif. Il est difficile de dire dans quelle mesure il s'agit de niveaux intrinsèques à la *description* du langage, indépendants de l'aspect applicatif. Il semble que leur cristallisation ne soit pas étrangère au développement du TAL, dans la mesure où chaque niveau peut recevoir une représentation symbolique par l'IA.

L'analyse morphologique des mots, fondée sur des dictionnaires, a été le premier niveau implémenté sur ordinateur, à l'aide de techniques symboliques (dictionnaires, listes d'inflexions, etc.). Les théories syntaxiques ont suivi peu après, puis les modèles sémantiques, fondés en général sur des formalismes logiques. Les connaissances pragmatiques ont fait l'objet de représentations par scénarios (*scripts*) ou cadres-type (*frames*). Plus récemment, des théories du discours utilisent aussi des formalismes logiques, implémentables.

La réduction symbolique a permis un nombre important d'avancées et de réalisations, même si on estime que certains objectifs du TAL sont assez loin d'être atteints. La vitalité de l'approche symbolique en TAL est manifestée par l'attribution du Prix d'Excellence pour la Recherche en IA, lors de l'*IJCAI-97*, à Aravind Joshi. Ses recherches en TAL suivent une démarche logico-symbolique :

> *L'ensemble d'idées qui caractérise [mes] efforts serait le mieux décrit comme l'utilisation de systèmes formels-computationnels contraints afin de décrire différents aspects du langage – syntaxe, sémantique, pragmatique et discours, parmi d'autres. […] L'utilisation de tels systèmes permet de distinguer entre les contraintes universelles [incorporées dans le système] et les contraintes normatives [ajoutées a posteriori car dépendant de la langue].* (Joshi 1998:95)

Nous verrons toutefois que les travaux de Joshi font appel aussi à des considérations non symboliques, notamment statistiques.

Les niveaux de description ou de traitement classiques sont relativement peu contestés. On peut certes remarquer une certaine interpénétration des niveaux : il y a par exemple une sémantique lexicale, et une sémantique de la phrase, leurs liens (compositionnalité de la première) faisant l'objet de nombreuses études. On a aussi étudié la syntaxe du discours, ou bien lexicalisé les règles syntaxiques. Les bases cérébrales de ces niveaux, dans les aires du langage, sont encore peu claires ; la distinction, fondée sur les aphasies, d'une aire plutôt lexicale (Wernicke) et d'une aire plutôt syntaxique (Broca) est assez grossière.

I.B.3 Réalisations du TAL symbolique

Un panorama du niveau actuel en TAL symbolique dépasse notre propos. On peut toutefois signaler le dynamisme des travaux en syntaxe formelle, ainsi que la construction de nombreuses ressources lexicales (morpho-syntaxiques et sémantiques) et syntaxiques (bases de règles, corpus étiquetés pour la syntaxe). L'accroissement des capacités informatiques de traitement et de stockage est en partie responsable de l'accroissement des performances, dont l'évaluation acquiert les dimensions et la rigueur d'un domaine de recherche en soi.

L'accroissement seul des capacités ne semble toutefois pas suffisant pour progresser. Ainsi, le projet CYC (Lenat et Guha 1990) vise depuis plusieurs années à accumuler des connaissances, exprimées symboliquement, sur le fonctionnement général du monde (*enCYClopédie*). Le projet utilise un moteur d'inférence pour exploiter l'immense base de connaissances, en vue de l'action rationnelle ou d'éventuelles découvertes ; toutefois, il ne semble pas avoir produit les résultats escomptés.

Examinons en revanche les trois exemples de systèmes formels défendus par Aravind Joshi (1998) :

a) les transducteurs à états finis, technique pour laquelle on constate un regain d'intérêt « grâce à notre connaissance substantielle du calcul à états finis ; grâce aux nouvelles techniques pour gérer d'énormes transducteurs… ; et, bien sûr, grâce aux techniques de gestion des transducteurs stochastiques et pondérés » ;

b) les grammaires d'arbres adjoints lexicalisées (LTAG), accompagnées de la technique du super-étiquetage des items lexicaux par des arbres. Le succès du super-étiquetage serait dû à la complémentarité des niveaux utilisés, ainsi qu'aux techniques statistiques ;

c) enfin, la théorie du centrage étudie la cohérence des énonciations d'un discours. De l'aveu de l'auteur, l'idée de la théorie est ancienne : « dans la représentation par calcul des prédicats, tous les arguments d'un prédicat ont le même statut. Cela n'est pas vrai pour une énonciation d'un discours. Il est intéressant de noter que, en 1879 déjà, Frege avait observé cette distinction et son utilité dans certains cas d'inférence ».

Ces contributions présentent donc deux points communs : tout d'abord, l'introduction de contraintes supplémentaires dans les modèles formels, qui proviennent de la prise en compte des niveaux supérieurs de structuration (interaction syntaxe/lexique, discours/phrase). D'autre part, l'usage de techniques non symboliques (ici statistiques), permettant d'orienter la résolution symbolique des problèmes. Nous allons donc examiner les principales techniques complémentaires proposées à ce jour.

I.C Nouvelles approches en intelligence artificielle et applications au traitement des langues

Un nombre croissant de travaux tentent de combler la distance entre les objectifs du TAL et son niveau actuel en utilisant des techniques différentes de celles de l'IA symbolique. La popularité et le succès de ces méthodes complémentaires sont assez variables, mais elles ont souvent le mérite de mettre en lumière de réelles limitations de l'IA symbolique. Toutefois, la manipulation de symboles demeure souvent un but à atteindre, mais sous une forme plus robuste et plus « intelligente ». Nous allons donner une série d'exemples où l'inspiration puisée dans d'autres domaines a permis des approches complémentaires de certains problèmes de l'IA, et en particulier du TAL, puis nous étudierons deux approches importantes : les réseaux de neurones et les systèmes multi-agents.

I.C.1 L'apport d'autres disciplines à l'IA

I.C.1.i La statistique

Nous venons de rencontrer les apports de la statistique dans les récentes avancées des techniques d'analyse syntaxique à états finis, et dans le super-étiquetage (I.B.3). La statistique a permis aussi l'amélioration du niveau des étiqueteurs morpho-syntaxiques, qui sont d'abord « entraînés » sur des corpus correctement étiquetés[10]. L'analyse syntaxique stochastique, utilisant des règles construites par apprentissage, est un sujet très actuel (Chappelier et Rajman 1998).

Un domaine ayant enregistré des progrès notables grâce aux statistiques est sans conteste le traitement ou la transcription de la parole, à l'origine une branche de l'IA. Des méthodes symboliques assez sophistiquées ont été employées, tels les tableaux noirs dans *Hearsay* (Erman et Lesser 1980). Ces approches ont été presque totalement supplantés par les modèles statistiques, fondés en général sur des *modèles de Markov cachés* ou *HMM* – voir par exemple (Waibel et Lee 1990). Ce changement s'est produit sous la pression de campagnes d'évaluation (cf. Chapitre VIII), les *HMM* conduisant à de meilleures performances mesurables.

I.C.1.ii Modèles continus vs. modèles discrets

Les théories physiques, ainsi que la théorie mathématique des systèmes dynamiques, ont été utilisées pour produire des modèles de compréhension du langage assez éloignés des modèles symboliques. Parmi les théories les plus exploitées, il y a les systèmes dynamiques, la thermodynamique et la théorie des catastrophes – voir des exemples dans (Perez 1988) et dans (Fuchs et Victorri 1994). Le modèle de construction

[10] Voir par exemple l'action française d'évaluation d'étiqueteurs nommée GRACE (Adda, Mariani, Lecomte, Paroubek et Rajman 1998).

du sens proposé par Bernard Victorri (1992, 1998) fait émerger la signification d'une phrase comme résultat de l'interaction dynamique de contraintes locales entre les opérateurs associés aux items lexicaux.

Les résultats de ces propositions sont pour l'heure plus difficiles à évaluer que ceux de la section précédente. On peut mentionner une discussion théorique (Kayser 1994) sur les modèles *continus*, la continuité étant une caractéristique marquante de ces propositions. Kayser cherche à montrer qu'il n'y a pas d'argument *a priori* pour ou contre le caractère continu ou discret des modèles ; il critique ainsi deux « preuves ontologiques » et une « preuve computationnelle »[11]. Kayser défend ici plutôt l'utilisation, pour le TAL, d'une multiplicité de représentations symboliques (donc discrètes), adaptées au niveau de détail de la tâche.

I.C.1.iii Algorithmes génétiques

L'inspiration biologique a donné naissance notamment à la technique des algorithmes génétiques, proposée par David Goldberg (1994). Il s'agit en fait d'une technique d'optimisation, qui s'applique particulièrement au traitement de l'information symbolique, comme le montrent Holland et Reitman (1978) dans un article traitant des « systèmes cognitifs fondés sur des algorithmes adaptatifs ».

Une façon de représenter le traitement de l'information dans ce formalisme est de considérer chaque « chromosome » comme une règle d'action, le début constituant les prémisses et la fin l'action (classifieurs génétiques). La mutation et la recombinaison des chromosomes, et leur sélection grâce à une fonction de récompense, permettent d'optimiser la réponse « moyenne » de l'ensemble, et obtenir un système en perpétuelle évolution adaptative. Cette utilisation des algorithmes génétiques demeure dans le paradigme de l'IA symbolique, enrichi d'une méthode d'optimisation.

D'autres utilisations des algorithmes génétiques s'écartent très sensiblement du paradigme symbolique, comme par exemples les études « neuro-génétiques », qui visent à faire évoluer un programme de construction de réseaux de neurones (détails au II.C.1.ii).

I.C.2 Autour des réseaux de neurones

L'idée d'une simulation informatique de l'architecture cérébrale et des processus qui s'y déroulent remonte au modèle du neurone proposé par McCulloch et Pitts (1943). Curieusement, cette autre possibilité théorique de réalisation d'une « intelligence artificielle » n'apparaît pas dans l'argumentation de Turing (1950). On peut rappeler la

[11] Il s'agit des trois raisonnements suivants, fallacieux selon Kayser : (1) le monde est (perçu comme) continu, le langage parle du (mode dont on perçoit le) monde, *donc* les modèles sémantiques doivent être continus ; (2) le langage est discret, les résultats demandés au TAL le sont aussi, *donc* les étapes intermédiaires de traitement doivent l'être aussi ; (3) les modèles discrets peuvent avoir une interprétation logique, la logique demandée par le TAL est par essence incomplète, *donc* les modèles discrets sont à éviter – d'après (Kayser 1994:112-114).

réalisation par Rosenblatt du « perceptron », un modèle simple contenant deux couches de neurones formels (entrée/sortie), reliés par des connexions de poids variable. Les limites de ce modèle furent dépassées au début des années 1980, avec l'introduction de couches cachées, et la découverte d'algorithmes d'apprentissage permettant de les utiliser. Nous évoquerons ici brièvement les possibilités de ce paradigme pour le TAL, et examinerons sa complémentarité par rapport aux techniques symboliques.

I.C.2.i Méthodes et applications

Les *réseaux de neurones artificiels* (RN) sont devenus un domaine de recherche en soi, séparé à la fois de la source d'inspiration (neurones naturels) et dans une certaine mesure même des objectifs généraux de l'IA. Les avancées du domaine semblent être dues surtout aux nouveaux algorithmes d'évolution des connexions d'un RN, souvent inspirés de la physique[12]. L'objectif générique est de traiter des données introduites dans la couche d'entrée, et produire une réponse jugée convenable par le programmeur. On classe souvent les RN selon l'algorithme d'apprentissage, qui peut être supervisé ou non (S/NS) ; aussi, l'information peut circuler uniquement de l'entrée vers la sortie, ou dans les deux sens (1/2). On obtient ainsi quatre catégories, avec des exemples : (S-1) perceptron (et son extension avec rétropropagation de gradient), (S-2) machine de Boltzmann, (NS-1) mémoire associative linéaire, (NS-2) modèles de Hopfield et de Kohonen.

Les applications des RN sont très variées ; un dénominateur commun est que ceux-ci permettent de classer facilement (selon des catégories prédéfinies ou non) de grands volumes de données. Cela recouvre aussi des tâches plus particulières comme la prédiction, la reconnaissance des formes, le contrôle, et de nombreuses modélisations. À peu près toutes les branches scientifiques et industrielles en utilisent les applications : statistique, physique, chimie, météorologie, économie, agriculture, médecine, biologie.

Il faut réserver une place à part aux RN utilisés pour la modélisation en neurosciences (dites computationnelles) et en sciences cognitives. C'est au fond le retour du modèle à sa source d'inspiration, et on constate que les algorithmes d'apprentissage classiques et les structures neuronales associées ne rencontrent pas toujours l'assentiment des neurobiologistes. Cela fait que les réalisations que ceux-ci souhaitent reposent sur des règles d'apprentissage et des architectures particulières aux phénomènes biologiques que l'on cherche à reproduire. On peut ainsi rappeler que, alors que les neurones artificiels sont en général de simples sommateurs à seuil, les neurones naturels ont des propriétés bien plus complexes : synapses de natures variées, dynamique temporelle , etc. Nous rencontrerons au chapitre suivant, avec les modèles de Gerald Edelman, de tels exemples de RN « biologiquement réalistes ».

[12] Pour une introduction au domaine on peut consulter (Fausett 1994), ou à un niveau plus avancé (Bishop 1995, Hertz, Krogh et Palmer 1991).

I.C.2.ii L'application à l'IA et au TAL : le connexionnisme symbolique

Une place à part doit être faite aux études cherchant à utiliser les RN pour répondre à certaines limitations de l'intelligence artificielle. Les propositions sont nombreuses, mais un certain courant semble se dessiner, participant à la fois de l'informatique, des sciences cognitives, des neurosciences et bien sûr de l'IA[13]. Il est à remarquer que ce courant ne remet pas en question la conception de l'intelligence comme capacité à manipuler les symboles ; au contraire, ces études visent à reproduire en un certain sens la fonction symbolique par les réseaux neuronaux, comme étant la voie vers des modèles cognitifs « de haut niveau »[14].

On a pu montrer d'un point de vue théorique qu'une machine de Turing pouvait être simulée par un RN, et inversement les RN sont en général programmés sur des ordinateurs symboliques[15]. Il ne s'agit donc pas de reproduire à l'aide d'un RN tel ou tel formalisme logico-computationnel, mais d'enrichir les systèmes symboliques avec des propriétés complémentaires. Les systèmes connexionnistes sont *subsymboliques*, à savoir ne manipulent pas directement les symboles, mais ceux-ci résultent d'une interprétation des entrées/sorties par des humains.

TRAITEMENT	SYMBOLIQUE (SPS) ou	SUB-SYMBOLIQUE
Contraintes	logiques	neurologiques
Connaissance	individualisée	distribuée
Types/instances	distincts	confondus
Mémoire	statique, structurée	dynamique, restructurable
Capacité de stockage	illimitée	fixée
Données/processus	séparés	confondus
Règles	déclaratives	émergeant statistiquement
Parallélisme	accessoire	essentiel
Catégories de base	prédéfinies	émergeant statistiquement
Résistance au bruit	fragile	robuste
Modèles	programmés	entraînés, renforcés
Apprentissage	preuves, inductions	adaptation incrémentale

Table I-1. Comparaison des approches symbolique et subsymbolique en intelligence artificielle – d'après (Dyer 1991:34).

[13] Ce courant est illustré par des recueils comme *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (McClelland, Rumelhart et The PDP Research Group 1986, Rumelhart et McClelland 1986b), ou encore *Connections and Symbols* (Pinker et Mehler 1988).

[14] Voir par exemple les recueils *High-Level Connectionist Models* (Barnden et Pollack 1991) et plus récemment *Connectionist Symbol Processing: Dead or Alive ?* (Jagota, Plate, Shastri et Sun 1998).

[15] Toutefois, de nombreux processeurs dédiés existent déjà, dont certains sont analogiques.

Dans une synthèse consacrée à l'approche subsymbolique (« neurosymbolisme ») appliquée au TAL, Michael Dyer (1991) propose une comparaison avec l'approche symbolique (le SPS, cf. I.A.2). Après avoir commenté les aspects pour lesquels ces méthodes sont comparables (cf. Table I-1), Dyer montre que celles-ci « habitent des niches computationnelles différentes », et souligne leur *complémentarité*. Ainsi, parmi les capacités de chacune, nous résumons :

- *pour l'approche symbolique* : utilisation de variables, valeurs et affectations ; structures récursives ; valeurs par défaut et héritage ; séparabilité (localisation) de la mémoire ; apprentissage explicite en une passe ; existence de langages de programmation séquentiels, par rapport auxquels la description actuelle des réseaux neuronaux ressemble au langage machine ;

- *pour l'approche subsymbolique* : intégration des connaissances ; continuité entre les représentations ; mémoire en permanente reconstruction ; auto-organisation ; associativité ; robustesse ; adaptabilité.

L'objectif des travaux de Dyer et de son équipe est de cumuler les avantages des deux approches, et nous en verrons un exemple au chapitre suivant (II.C.3.ii).

Parmi les réalisations en TAL connexionniste (ou subsymbolique), on rencontre souvent des modèles à visée psycholinguistique, qui tentent de rendre compte de faits observés. On peut citer des études sur l'acquisition de la morpho-syntaxe (Prasada et Pinker 1993, Rumelhart et McClelland 1986a), et le modèle de (McClelland et Kawamoto 1986) qui étudie l'analyse syntaxique au niveau de la phrase. En ce qui concerne la sémantique, on peut citer le modèle de la polysémie de (Victorri et Raysz 1989), et l'étude de l'émergence de catégories sémantiques de (Elman 1990).

L'utilisation réussie des RN en reconnaissance des formes suggère que ceux-ci sont une bonne solution aux problèmes de la perception artificielle. Or, un des problèmes du TAL symbolique est justement l'absence de fondement perceptif pour le langage manipulé. Dyer (1991:75) résume ainsi ce problème :

> *Les systèmes [actuels] de TAL ne possèdent pas des représentations qui relient les mots à l'information sensorielle provenant/traitant du monde. Les modèles symboliques aussi bien que les modèles PDP[16] destinés au TAL possèdent dès le départ des représentations des rôles syntaxiques/sémantiques (p.ex. SUJET, ACTEUR, OBJET, MODIFIEUR) et des relations (p.ex. CAUSE, POSSÈDE, ÉPOUSE) […] Mais il n'y a pas de correspondance dans ces modèles entre les structures conceptuelles internes et l'expérience sensorielle du monde. La difficulté d'ancrer le langage dans l'expérience sensorielle a rendu difficile la modélisation de l'apprentissage du langage.*

[16] « Parallel Distributed Processing », i.e. à base de réseaux de neurones (McClelland, Rumelhart et The PDP Research Group 1986, Rumelhart et McClelland 1986b).

Une bonne intégration entre la perception artificielle et le TAL, par le biais des techniques neurosymboliques, permettrait donc de répondre à ce problème.

I.C.3 Agents et systèmes multi-agents

La modélisation *multi-agent*, et la notion reliée d'*agent*, recouvrent un spectre très large de travaux (Ferber 1995). L'idée de ce courant est de répartir les tâches à exécuter entre plusieurs modules plus ou moins autonomes, capables d'interagir, et *a fortiori* situés dans un environnement plus ou moins riche. Le succès de l'approche est dû au fait, constaté expérimentalement, que le comportement collectif des agents (même rudimentaires) manifeste une complexité et une capacité de traitement bien supérieures à celles de chaque individu. On est ainsi en présence d'une *intelligence collective* qui peut paraître une solution de remplacement à l'IA symbolique : tel est par l'exemple le message de Rodney Brooks dans ses deux articles (nous traduisons) *Les éléphants ne jouent pas aux échecs* et *Intelligence sans représentation* (Brooks 1990, 1991).

Cette autre façon de *distribuer* le traitement (après le connexionnisme) a vu aussi de très nombreuses applications, qu'elles soient théoriques (algorithmes multi-agent), technologiques (systèmes de contrôle, d'optimisation, de gestion de l'information) ou pour la modélisation (modèles de sociétés en tout genre). Son application au langage naturel peut *a priori* emprunter deux directions : la capacité linguistique peut émerger au niveau du système, ou bien à appartenir à chaque agent.

La première possibilité suppose en général une répartition des tâches et des connaissances linguistiques entre agents, et la définition d'un mécanisme de coopération permettant l'analyse linguistique. On peut citer ici les modèles de compréhension CARAMEL (Sabah 1996, Sabah et Briffault 1993), fondé sur une extension des tableaux noirs et TALISMAN (Stéfanini et Demazeau 1995), ou encore un système multi-agent pour la résolution des anaphores pronominales (Rich et LuperFoy 1988). Les « agents » utilisés dans cette approche sont doués d'une autonomie assez faible, et leurs « intentions » sont rarement explicitées. La difficulté principale nous paraît être ici la distribution de l'expertise entre les différents agents, et la définition de structures de données partageables (Popescu-Belis 1996).

La seconde possibilité utilise des agents doués eux-mêmes de capacités linguistiques autonomes, dont l'interaction entraîne l'émergence d'un code de communication collectif, fondé sur des conventions décentralisées et évolutives. Il s'agit donc de *simulations* de phénomènes langagiers, plutôt que de systèmes orientés directement vers le *traitement* de la langue, du moins dans un premier temps. Cette approche permet de saisir le caractère collectif, dynamique et conventionnel du langage, et permet de répondre au problème de l'ancrage perceptif du langage, par l'utilisation d'agents situés dans un environnement simulé, qu'ils peuvent percevoir, et aussi traiter en fonction de leurs besoins – nous y reviendrons en détail au II.D.2. La difficulté consiste ici à munir un agent de capacités linguistiques, ce qui ne ferait que reporter le problème du TAL ; ce problème semble en tout cas mieux posé pour un *agent situé* que pour un programme symbolique.

I.D Notre proposition : la « pragmatique simulée »

Ce passage en revue des différentes propositions complémentaires au paradigme de l'IA symbolique met en évidence, à nos yeux, un dénominateur commun : ces démarches préconisent toutes une reproduction plus fidèle des *mécanismes* du langage afin de pouvoir traiter celui-ci, i.e. simuler la *fonction* (cf. note 1). Chacune des propositions met l'accent sur un aspect différent du mécanisme réel du langage, qui reste pour l'heure assez mal connu, en particulier dans ses bases cérébrales. Nos analyses vont nous permettre de justifier nos choix de mécanismes : ceux qui, à la fois, paraissent prioritaires pour la compréhension de la langue par l'ordinateur, et en même temps commencent à être à la portée des techniques d'implémentation actuelles.

Nos propositions se développent ainsi selon plusieurs lignes de force. Tout d'abord, nous cherchons à mieux reproduire le *caractère situé* du langage humain – à savoir son ancrage dans l'*individu*, sa relation à *l'environnement* grâce aux perceptions et actions de l'individu, et son fondement *collectif*, à travers les conventions linguistiques et le rôle communicatif du langage. La nécessité de simuler des individus et un environnement conduit à l'utilisation de la *modélisation multi-agent* ; la difficulté de fournir aux agents de nombreuses connaissances linguistiques, ainsi que la modélisation de leur caractère conventionnel, conduisent à une approche fondée sur *l'émergence* de conventions linguistiques.

À l'intérieur de cet axe de recherche, nous défendons de façon plus originale l'importance du lien entre langage et environnement, par le biais des représentations. Des expressions du langage permettent en effet d'établir ou de signaler le lien entre des entités de l'environnement et des représentations évoquées dans le discours. Le phénomène de la *référence* est en effet essentiel, et nous tenterons de montrer comment il peut être pris en compte dans une modélisation multi-agent du langage. Faisant écho aux théories de la pragmatique (Moeschler et Reboul 1994), i.e. à l'étude de la communication dans son contexte concret, nous proposons d'appeler « *pragmatique simulée* » l'approche développée ici, puisqu'elle vise à recréer par la simulation la possibilité d'une communication véritablement *pertinente* entre deux agents rationnels.

On comprend donc que notre thèse fasse appel à une grande variété de disciplines ayant trait à l'étude du langage et à l'intelligence artificielle, ainsi qu'à des théories neurologiques, psychologiques et linguistiques abordant le mécanisme de la langue en général et de la référence en particulier.

Les considérations sur l'évaluation des réalisations, là où celle-ci est possible, permettent d'enregistrer les progrès réalisés, et justifier les techniques employées. Nous montrerons à ce niveau qu'un des modèles les plus pertinents, parmi les mesures d'évaluation étudiées, fait intervenir la transmission d'*information référentielle* entre deux agents (source et récepteur) munis de représentations des référents.

Chapitre II

RÉALITÉS ET SIMULATIONS DE LA COMMUNICATION LANGAGIÈRE

Nous avons présenté dans le chapitre précédent un échantillon varié d'approches adoptées en intelligence artificielle et en traitement automatique de la langue. Afin de montrer sur quels points celles-ci doivent être étendues, grâce à la « pragmatique simulée », nous allons mettre en évidence quelques éléments clés du langage qui éclairent sa simulation informatique.

Nous examinerons d'abord les *réalités* de la langue, décrites par les sciences expérimentales : neurosciences, psychologie, linguistique descriptive. Nous porterons un intérêt particulier aux études des bases psychophysiques de la langue, et des différents cas d'émergence. On peut distinguer deux plans : celui de l'*individu* (II.A), où nous rappellerons les connaissances consensuelles (II.A.1) avant de décrire plus en détail les propositions de Gerald Edelman (II.A.2). Pour la *collectivité* (II.B), nous dresserons un tableau des trois cas d'émergence du langage connus à ce jour : l'évolution du langage dans l'espèce (II.B.2), son acquisition par les enfants (II.B.3), et sa réinvention dans les communautés créoles (II.B.4). Ces cas ont un intérêt particulier pour les modèles de l'émergence du langage (II.B.1).

La seconde partie du chapitre s'intéresse aux modèles informatiques de la communication linguistique. Nous commencerons aussi par des modèles d'*individus* isolés ou *agents* (II.C), plus ou moins autonomes ou doués de capacités linguistiques (II.C.1 à II.C.3), en particulier les modèles élaborés dans l'équipe d'Edelman (II.C.2). Ensuite, nous passerons en revue les modèles de *collectivités d'agents* communicants (II.D), en privilégiant les simulations d'échanges linguistiques (II.D.2) par rapport aux protocoles de communication formels (II.D.1).

II.A Bases individuelles du langage

Il faut distinguer trois démarches traditionnelles complémentaires dans l'étude du phénomène de la langue. La *linguistique* se concentre sur les propriétés intrinsèques de la langue, souvent considérée comme un système autonome. Les *neurosciences* étudient les bases cérébrales du langage, entre autres facultés. La *psychologie cognitive* vise à saisir et analyser de la façon la plus précise les manifestations de la capacité linguistique, parmi d'autres capacités de niveau supérieur. Les trois disciplines ont vu l'éclosion de nombreux modèles fondés sur les observations expérimentales. Signalons aussi qu'il

existe de nombreuses études à l'interface des trois domaines (neurolinguistique, psycholinguistique, neuropsychologie du langage, sciences cognitives).

Les descriptions de la faculté de langage intéressent au plus haut point le chercheur en intelligence artificielle et traitement automatique des langues. En effet, une meilleure compréhension du mécanisme ne peut qu'aider à sa simulation. Se pose donc la question suivante, résumée ici par Noam Chomsky (1995:1-2) :

> *[Il y a] un problème pour la biologie qui est loin d'être trivial : comment un système tel que le langage humain peut apparaître dans l'esprit ou le cerveau, ou plus généralement dans le monde organique, là où on ne trouve rien qui ressemble aux propriétés élémentaires du langage ? Ce problème a parfois été vu comme une crise pour les sciences cognitives. […] La biologie et les sciences du cerveau, dans l'état actuel, ne fournissent aucun fondement pour des propriétés du langage qui sont bien établies [en linguistique].*
> *Pour une discussion de cette question, voir (Chomsky 1994), qui fait référence à (Edelman 1992a). Edelman estime que la crise est sérieuse, voire fatale, pour les sciences cognitives en général, qu'elles soient computationnelles, connexionnistes, ou autres.*

Afin de mesurer l'étendue de cette crise, nous allons résumer ici les principaux éléments apportés par les sciences expérimentales à l'étude du langage, avant d'examiner les propositions plus novatrices de Gerald Edelman. Nous restons dans une perspective « synchronique » de la langue, et aborderons les questions de développement dans la section suivante (II.B).

II.A.1 Connaissance des mécanismes du langage : neurosciences, psychologie, linguistique

II.A.1.i Bases cérébrales du langage

La connaissance du cerveau a progressé à la fois grâce aux nombreuses études moléculaires et cellulaires, et grâce à l'expérimentation animale, en particulier *in vivo*. Ces techniques, qui dévoilent la structure des neurones, les mécanismes synaptiques ou l'organisation des aires perceptives par exemple, ne sont que d'une utilité limitée pour l'étude du langage. Les deux méthodes d'investigation clé pour ce sujet sont l'étude de des lésions cérébrales et plus récemment l'imagerie[1]. Toutefois, l'apport des neurosciences à la modélisation informatique reste encore assez indirect. Tout au plus demandera-t-on aux modèles informatiques « réalistes » de ne pas contredire les résultats que nous allons synthétiser ici.

[1] De très nombreux livres abordent ce sujet. Nous utilisons l'important livre de Kandel *et al.* (1995), qui présente un niveau de détail suffisant pour un informaticien, et une ouverture vers les neurosciences cognitives. Nous nous limitons bien sûr aux conclusions qui font l'accord des chercheurs ; l'étude des travaux encore débattus dépasse notre but.

La plupart des descriptions neurologiques des bases du langage consacrent une large part aux liens entre les aphasies (troubles du langage) et les lésions cérébrales qui les provoquent. Les travaux de Broca et Wernicke au siècle dernier avaient déjà permis de localiser deux aires de l'hémisphère gauche dont la lésion causait des aphasies ; leurs continuateurs ont peu à peu affiné la carte des « aires du langage » et de leurs connexions, et leurs liens avec les déficits respectifs. L'imagerie cérébrale mesure l'activité des aires corticales lors de l'exécution de diverses tâches linguistiques ; sa résolution spatiale et temporelle est encore à améliorer.

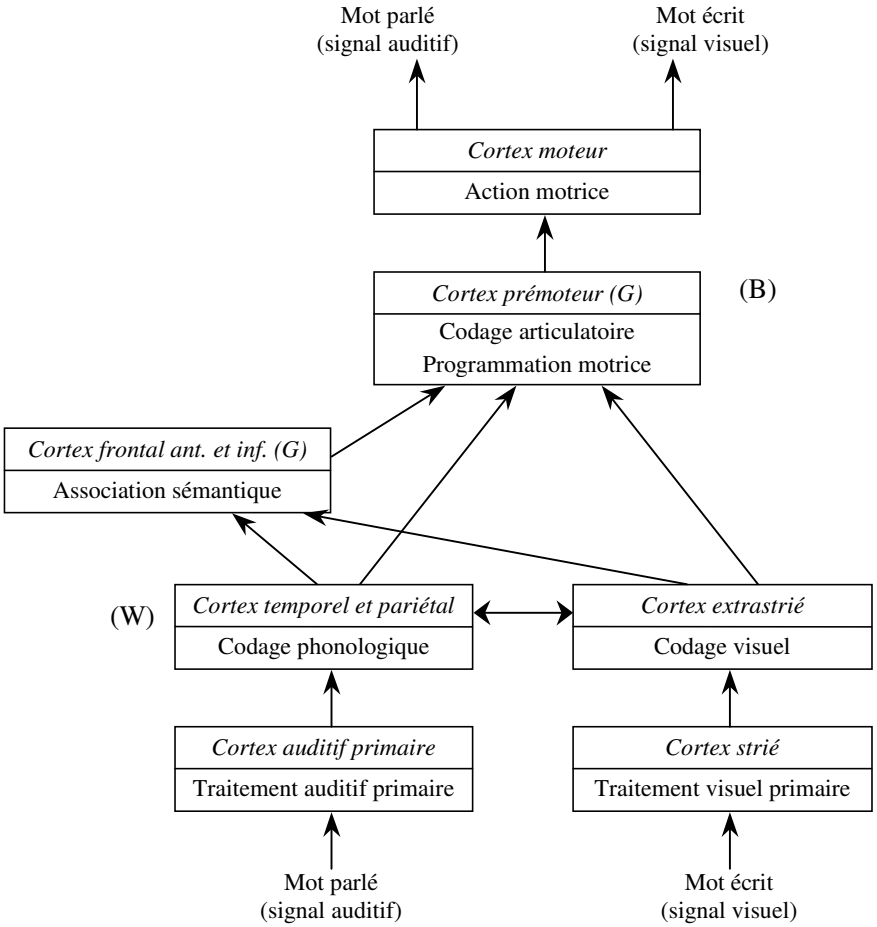


Figure II-1. Schéma simplifié des aires corticales et connexions intervenant dans le traitement du langage – (Kandel, Schwartz et Jessell 1995:645) d'après (Petersen, Fox, Posner, Mintun et Raichle 1989). Il s'agit d'une mise à jour consensuelle du modèle de Wernicke-Geschwind. L'aire (B) est affectée dans l'aphasie de Broca, et l'aire (W) dans celle de Wernicke, et on leur attribue souvent ces noms.

Les acquis incontestables sont résumés sur la Figure II-1. On remarque que le traitement linguistique proprement dit reste largement inexploré : on l'attribue souvent au « cortex associatif ». Des propositions existent pour le statut du lexique, utilisant justement les associations entre stimuli ; d'autres travaux explorent le rôle du cortex préfrontal dans l'utilisation de la syntaxe, car cette région joue un rôle dans la

« planification »[2]. Toutefois, la microstructure du cortex associatif, et donc le codage neuronal des différents éléments linguistiques restent largement inconnus.

II.A.1.ii Études psychologiques du langage

Les études psychologiques réalisent une exploration non intrusive, comportementale, des sujets humains, à l'aide d'expériences rigoureuses et reproductibles. L'avènement de l'imagerie cérébrale a rendu plus floue la frontière entre psychologie cognitive et neurosciences. Ces dernières s'intéressent à des modèles plutôt *structurels* de la faculté de langage, alors que les psychologues proposent le plus souvent des modèles *fonctionnels*. L'étude psychologique des mécanismes du langage est rendue difficile par l'accès limité qu'en donnent la parole et le langage écrit, qui constituent seulement l'interface d'entrée/sortie de la faculté de langage.

De nombreuses études sont consacrées à la perception du langage écrit, cf. par exemple (Nazir sous presse), et surtout de la parole, cf. par exemple (Segui 1989). Le lien entre langue et représentation spatiale est étudié au LIMSI par Michel Denis (Denis et Vega 1993), Xavier Briffault (Briffault 1995), Gérard Ligozat (Ligozat 1994). Un modèle psychologique (fonctionnel) de la planification figure dans (Shallice 1988) ; la planification a été reliée à la syntaxe dans les études de A. Sirigu *et al.* (Thierry, Glowinski, Goldman-Rakic et Christen 1994). Le rôle de la mémoire dans la compréhension des phrases est examiné dans la récente synthèse de D. Caplan et G. Waters (Caplan et Waters à paraître), qui montre que la mémoire intervenant dans la compréhension « syntaxique et sémantique » est distincte de la « mémoire à court terme » utilisée à des fins non linguistiques. Une question étudiée avec beaucoup de succès est le développement cognitif et linguistique de l'enfant (cf. II.B.3).

Relativement peu d'études peuvent être exploitées directement pour la modélisation informatique. On doit en effet interpréter les conditions expérimentales (le protocole) puis les résultats eux-mêmes. Les modèles intermédiaires proposés par les psychologues semblent, par leur caractère fonctionnel, moins applicables que ceux des neurologues. L'informaticien utilisera alors plutôt des théories neuro-psychologiques déjà constituées, comme celle de Gerald Edelman (cf. II.A.2).

II.A.1.iii Un organe du langage ?

La faculté de langage est fondée sur une certaine structure cérébrale, fruit entre autres d'un processus de développement complexe, et de l'interaction entre individus. Le débat sur la nature et le rôle exact de ces éléments est souvent ramené à une opposition déclinée sur trois axes majeurs :

a) Quel est le statut de la faculté de langage par rapport aux autres facultés supérieures de l'entendement ? S'agit-il de capacités distinctes, ou bien l'usage d'une langue

[2] Voir par exemple certains articles du recueil de (Thierry, Glowinski, Goldman-Rakic et Christen 1994) sur le cortex préfrontal.

relève-t-il des mêmes dispositions que la manipulation des nombres, ou l'orientation dans une ville ?

b) Y a-t-il dans le cerveau des aires exclusivement dédiées au traitement du langage, dont la structure et les connexions sont particulières à cette fonction, ou bien les fonctions supérieures sont-elles uniformément réparties dans le cortex associatif ?

c) Enfin, la capacité linguistique est-elle innée dans chaque individu, ou bien est-elle entièrement acquise durant l'exposition au langage, en même temps que les autres aptitudes transmises culturellement ?

Parmi les débats classiques couvrant un ou plusieurs de ces points, il faut citer celui entre Chomsky et Piaget (Piatelli-Palmarini 1976). Plus récemment, on peut opposer deux livres ayant rencontré un succès notable : celui de S. Pinker (1994) intitulé *L'instinct du langage*, et celui d'un groupe de chercheurs de San Diego (Elman, Bates, Johnson, Karmiloff-Smith, Parisi et Plunkett 1996) intitulé *Repenser l'innéisme*.

Un des principaux « observables » du problème est le processus d'acquisition du langage par l'enfant (cf. II.B.3), où l'on tente de discerner ce qui relève d'un « mécanisme biologique programmé » et ce qui relève d'une « transmission culturelle » des conventions linguistiques (question *c*). Très utilisées aussi sont les études des déficits cognitifs et langagiers chez les patients : les troubles du langage sans trouble cognitif global sont bien attestés (aphasies), alors que le déficit inverse semble plus rare[3] (questions *a* et *b*). En réalité, ces oppositions sont quelque peu artificielles. Le débat s'oriente en réalité vers la *proportion* de chaque élément – « inné » resp. « acquis » – dans le phénomène du langage.

Pour l'informaticien, l'existence d'un « organe du langage » très spécifique remettrait en question la visée de l'intelligence artificielle *symbolique*, qui vise plutôt à reproduire les « capacités cognitives générales ». Par leurs modèles, les informaticiens contribuent aussi à valider ou infirmer les suppositions de ce débat.

II.A.1.iv Fondements de la capacité linguistique : travaux de Chomsky

Parmi les modèles de la faculté de langage proposés par des *linguistes*, nous esquissons ici, à titre d'exemple important et développé, la *Théorie des Principes et Paramètres (P&P)*. Fruit des travaux de Noam Chomsky et de ses collaborateurs (Chomsky 1995, Chomsky et Lasnik 1993), cette théorie prend la relève de la « grammaire générative », et fait encore l'objet de nombreux débats.

Le modèle postule l'existence d'une *description structurelle* (SD), qui est un « objet symbolique […] représentant l'ensemble des propriétés phonétiques, sémantiques et syntaxiques d'une expression linguistique particulière » (Chomsky 1995:15) – il s'agit donc d'un objet mental et formel à la fois. Spécifier une langue revient à spécifier un ensemble (infini) de SD autorisées. La faculté de langage consiste en un système

[3] Ce serait le cas dans le syndrome de Williams – cf. (Pinker 1994:50-53) et ses références.

computationnel, qui spécifie les SD, et un système lexical, qui spécifie les items lexicaux figurant dans la SD. Lorsque ces items ont été insérés dans la SD, la représentation obtenue est appelée *D-Structure* (à l'origine, *structure profonde*).

La SD est utilisée par d'autres systèmes, dits de « performance », qui possèdent chacun un niveau de représentation différent. Ainsi, on dérive à partir de la SD à la fois la *forme phonétique* (PF) de l'expression, ainsi que sa *forme logique* (LF) qui est utilisée dans les opérations cognitives. Il est plus facile de caractériser le module phonétique, plus accessible, que le module logique, qui est « plus obscur : contient-il un système conceptuel, une théorie pragmatique du sens, une théorie causale de la référence ? » (Chomsky 1995:21).

La dérivation de PF et LF à partir de la D-Structure suit d'abord un même processus, jusqu'à la S-Structure, puis se bifurque (Figure II-2). La D-Structure se projette donc sur la S-Structure, qui seule se projette en PF et LF.

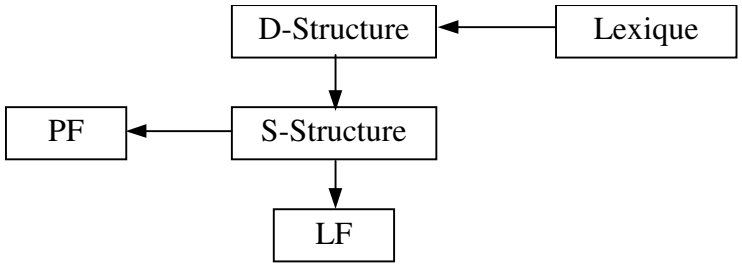


Figure II-2. Objets mentaux linguistiques selon (Chomsky 1995:22).

La théorie P&P affirme que les seuls éléments particuliers à une langue sont les propriétés du lexique, et le niveau où se séparent les branches pour PF et LF sur la transformation de la D-Structure. Les autres mécanismes ne varient pas selon les langues, en particulier les interfaces de PF et LF avec les autres parties du système cognitif. Le but principal de la théorie est de formuler à la fois les principes universels et les paramètres, et d'étudier pour chaque langue les valeurs des paramètres.

L'aperçu donné ici suffit pour comprendre que les bases conceptuelles de la théorie P&P ne semblent pas fondées sur des considérations neuropsychologiques, mais proviennent d'une réflexion abstraite sur des manifestations du langage détachées de leur contexte. Le formalisme linguistique utilisé permet certes d'établir des propriétés linguistiques corroborées par l'expérience, et même de construire des programmes d'analyse syntaxique (Berwick, Abney et Tenny 1991). Mais cela ne suffit pas pour accréditer le modèle, auquel on peut reprocher aussi le flou qui entoure l'interface logique (le « sens »). Il ne faut pas oublier qu'il s'agit d'*une théorie syntaxique*, pour laquelle l'*ancrage* du langage dans le monde n'est pas central.

II.A.2 La théorie de Gerald M. Edelman : du « darwinisme neuronal » au langage et à la conscience

Nous abordons maintenant une théorie essentielle pour notre approche. Cette théorie a le mérite de partir de considérations sur la microstructure du cerveau, pour aboutir à une description fonctionnelle des facultés supérieures, fondées dans la macro-structure du cerveau. Deux livres traitent de la microstructure, *Topobiology* (Edelman 1988) et surtout *Neural Darwinism* (Edelman 1987). Le premier, très technique, s'intéresse aux facteurs biochimiques de la croissance des cellules et des connexions formant le système nerveux ; le second développe la Théorie de Sélection des Groupes de Neurones (TSGN) – une théorie de l'évolution microscopique du système nerveux avant et après la naissance de l'individu, qui explique son organisation par un mécanisme sélectionniste. Un troisième livre, *The Remembered Present* (Edelman 1989), applique ces idées à la macro-structure du cerveau, et fournit une tentative d'explication des fonctions supérieures comme la conscience, le langage, l'attention. Enfin, une synthèse de ces trois livres, accompagnée de considérations plus spéculatives, est donnée dans *Bright Air, Brilliant Fire : on the Matter of Mind* (Edelman 1992b)[4].

Afin de donner un aperçu de cette *théorie intégrée* qui vise, parmi d'autres buts, à expliquer la faculté de langage, nous irons de l'analyse *microscopique* vers les mécanismes *macroscopiques*. Les premières considérations peuvent donc sembler éloignées de la fonction langagière elle-même, mais serviront aussi à comprendre les modèles informatiques présentés plus loin (II.C.2). Étant donné l'extrême étendue de la théorie, nous ne pourrons rester qu'à un niveau très élémentaire.

II.A.2.i Architecture neuronale et catégorisation

II.A.2.i.a Groupes de neurones, répertoires, cartes

L'unité de sélection du darwinisme neuronal est le *groupe de neurones* (abrégé GN). Ces groupes possèdent une entrée et une sortie, et peuvent établir des connexions dites *réentrantes* (réciproques) avec d'autres groupes, ce qui n'est pas le cas des neurones isolés. Ceux-ci sont beaucoup plus soumis à la variation, et sont donc moins aptes à la structuration ; il serait irréaliste d'associer une fonction déterminée à un seul neurone. En revanche, on peut identifier des groupes jouant un rôle fonctionnel, par exemple les colonnes corticales détectant certaines propriétés locales du champ visuel[5].

---

[4] Les traductions en français des titres sont respectivement : *Darwinisme neuronal : la théorie de sélection des groupes de neurones – Topobiologie – Le présent remémoré : une théorie biologique de la conscience – Souffle clair, feu étincelant : sur la matière de l'esprit*. Seul ce dernier livre a été traduit en français, sous le titre *Biologie de la conscience*, un peu trompeur dans la mesure où il reprend le sous titre du troisième. Le titre original provient d'une citation d'Empédocle donnée en exergue, que nous identifions avec le fragment B, *cix* (in Aristote, *De anima*, I, *ii*, 404b8).

[5] Les micromodules que sont les colonnes corticales sont bien connus non seulement dans le cortex sensoriel (Kandel, Schwartz et Jessell 1995:440-444), et visuel en particulier (travaux de Hubel et Wiesel,

Les GN forment des structures nommées *répertoires*. Résultat des sélections synaptiques décrites par la TSGN, les répertoires sont matérialisés par un certain ensemble de connexions, plutôt que par un emplacement précis des GN. Les *cartes* constituent un cas particulier de répertoire[6] : ce sont des ensembles de GN qui sont réceptifs à certains patrons d'activité d'une région neuronale qu'ils *cartographient*. Ces cartes sont maintenant bien connues pour les premiers niveaux d'entrée sensoriels (surtout visuels), et Edelman suppose leur omniprésence dans le cortex.

Ces cartes permettent donc de représenter certains traits de l'activité d'un autre répertoire, grâce à l'activation d'un ou plusieurs GN pour une configuration spécifique du répertoire cartographié, qui peut être lui-même une carte. La succession des cartes permet une catégorisation de plus en plus abstraite des zones d'entrée. Dans une carte, plusieurs combinaisons de GN actifs peuvent correspondre au même stimulus et réciproquement un GN peut participer à plusieurs combinaisons, ce qui permet les généralisations opérées à l'aide des *couples de classification* (cf. II.A.2.i.c).

II.A.2.i.b La réentrance

La *réentrance* dénote l'échange synchrone de signaux entre deux cartes ; les connexions réentrantes recouvrent une grande variété morphologique (Edelman 1989:66). Nous distinguerons deux types de réentrance, « ascensionnelle » et « transversale »[7]. La première permet de renvoyer l'information élaborée dans une carte de niveau supérieur vers une carte de niveau inférieur, plus proche d'un niveau d'entrée perceptif, ce qui permet d'améliorer les performances de catégorisation, et même de faire apparaître des caractéristiques nouvelles dans la carte de niveau inférieur. La deuxième relie deux cartes de même niveau, par exemple dans deux modalités différentes. La réentrance synchronise et enrichit la catégorisation de chaque carte. Celles-ci forment alors un *couple de classification* (cf. II.A.2.i.c).

La réentrance ne doit pas être confondue avec une simple rétroaction (feed-back), comme certains l'ont reproché (Calvin 1988). Elle apporte des informations nouvelles, et

---

(1979)), mais plus généralement dans l'ensemble du néo-cortex (Kandel, Schwartz et Jessell 1995:351). Ces colonnes, d'une taille variant entre 30 et 500 µm, peuvent contenir plusieurs centaines de neurones ; leur rôle a été étudié aussi par Vernon Mouncastle (Edelman et Mountcastle 1978). Un modèle computationnel en a été proposé par Yves Burnod (1988).

[6] La distinction n'est pas très claire : les termes ne se mélangent jamais dans (Edelman 1989), et aucun n'est défini par rapport à l'autre.

[7] Chez Edelman, cette distinction est implicite, aucun nom n'étant donné. Mais cette distinction importante se retrouve dans d'autres travaux (Philips et Singer 1997).

conduit à un changement qualitatif lorsqu'elle procède de « mémoire des valeurs-catégories » vers les couches perceptives primaires et secondaires, puisqu'alors sont créées les conditions d'apparition de la conscience primaire. La conscience d'ordre supérieur, liée au langage, est due aussi à des connexions réentrantes. Il reste toutefois que d'un point de vue strictement informationnel la réentrance est bien une variété de rétroaction.

II.A.2.i.c Les couples de classification

La notion de couple de classification, assez technique, est introduite pour expliquer la catégorisation et la généralisation multimodales. Il s'agit de deux cartes, chacune pouvant détecter des traits dans une modalité sensorielle – elles sont soit connectées directement à une couche sensorielle, soit se trouvent à un niveau plus élevé. Chaque carte, ensemble de GN sensibles à certains traits des stimuli, remplit d'abord sa fonction de catégorisation des stimuli reçus.

Un apprentissage associatif de type hebbien, transversal et bidirectionnel, peut créer des liens réentrants entre les cartes. Ainsi, la finesse de la catégorisation aussi bien que le pouvoir de généralisation augmentent. La détection de certains traits par une carte peut contribuer à évoquer les traits qui leur sont associés dans l'autre carte, comme l'exemplifient aussi les automates *Darwin II-III* (cf. II.C.2.i).

II.A.2.i.d Quelques difficultés

Les considérations précédentes reposent souvent sur des données expérimentales, qu'elles prolongent ou et parfois dépassent largement. La notion de réentrance semble relativement bien fondée, cependant le seul exemple de fonctionnement donné est le couple de classification, où la réentrance intervient dans la généralisation. Invoquée très souvent, elle constitue la base d'un apprentissage associatif, comme l'interprétation que nous en donnerons dans notre simulation (cf. IV.A, surtout IV.A.3.iii)[8].

II.A.2.ii La Théorie de Sélection des Groupes de Neurones (TSGN)

La TSGN proprement dite décrit l'apparition des structures neuronales que nous venons de décrire, structures permettant la catégorisation, la généralisation et

[8] Un concept plus vague, non approfondi ici, est celui de *cartographie/correspondance globale* (en anglais « global mapping »). Cette difficile notion semble désigner un ensemble complexe de connexions réentrantes qui unissent les entrées aux zones de contrôle et aux sorties motrices du système nerveux. Un seul un exemple en est donné (Edelman 1989:55), mais cet exemple est purement graphique ; la notion est pourtant invoquée de nombreuses fois dans la suite, pour expliquer des propriétés de haut niveau. Ce manque de clarté nuit bien sûr à la réalisation de modèles computationnels, mais il semble que l'état actuel des connaissances on ne puisse pas être plus précis.

l'apprentissage. Selon la TSGN, trois grandes étapes sont à distinguer dans la genèse de ces structures :

- La sélection durant le développement somatique, au cours duquel se forment les répertoires primaires, grâce à des modifications structurales et topologiques. Il s'agit de la croissance même des neurones et de leurs ramifications, due essentiellement aux variations locales d'expression de certaines molécules (Edelman 1988). Cette étape a lieu pendant l'embryogenèse**.**
- La sélection par l'expérience, qui voit le renforcement ou l'affaiblissement de certaines synapses. Cela est le résultat de la présentation répétée de grandes classes de stimuli et des réactions développées face à ceux-ci, à travers des corrélations statistiques de signaux. Se forment ainsi les *répertoires secondaires*.
- La formation de cartes où se projettent les différentes couches sensorielles, par sélection entre de nombreux groupes neuronaux isofonctionnels, qui explique la variabilité phénotypique individuelle. Entre les cartes s'établissent des connexions réentrantes, qui leur permettent de corréler dans le temps leurs activités, d'échanger des informations concernant les stimuli, et former ainsi des n-uplets classifieurs. Cette formation est dite *épigénétique*.

Il ne ressort pas clairement si les étapes (b) et (c) ont lieu lors de chaque apprentissage, ou bien seulement au début de la vie, lorsque les grandes catégories perceptives se forment. On sait par exemple qu'il y a à cette époque certaines phases critiques pour la vision, par exemple, où une privation relativement courte entraîne des incapacités définitives. En revanche, à la maturité, l'apprentissage se révèle beaucoup plus robuste ; la TSGN ne semble pas préciser s'il s'agit du même mécanisme. En tout cas, les notions de répertoire primaire et secondaire sont liées aux étapes de formation du répertoire, et ne semblent pas servir à décrire l'état adulte – *on peut donc en faire l'économie dans une modélisation informatique, en construisant directement une structure de cartes*.

II.A.2.iii La conscience primaire et les concepts

Nous avons jusqu'ici mis en place les concepts microscopiques de la théorie d'Edelman, nécessaires pour comprendre les propriétés de niveau supérieur. Edelman introduit celles-ci en deux étapes, correspondant aussi à deux stades de l'évolution des espèces. La *conscience primaire* est liée à la capacité de manipuler des « concepts » (au sens edelmanien exposé ici), puis la *conscience de haut niveau* est étroitement liée au langage. Les indications d'Edelman se font ici plus schématiques, et aucun modèle computationnel ne vient les étayer. Son exposé, comme notre aperçu, repose beaucoup sur des schémas fonctionnels, compatibles avec les connaissances actuelles sur la structure du cerveau (voir aussi l'analyse de Gérard Sabah (1995:21-27)).

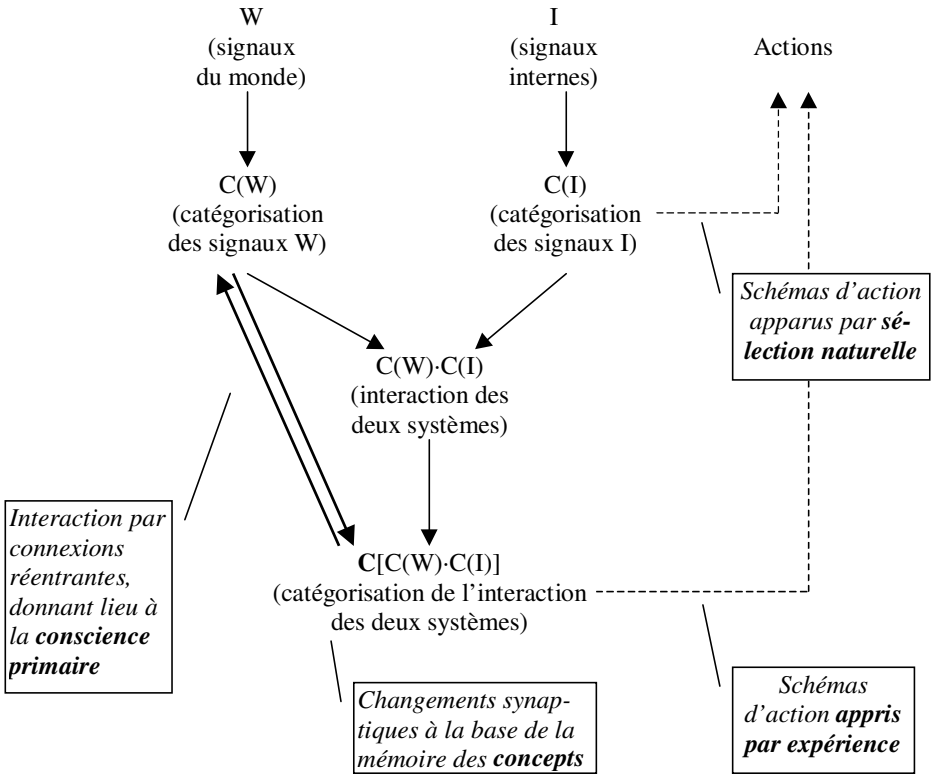


Figure II-3. La catégorisation intéro- et extéroceptive permet l'apparition des concepts, de la conscience primaire, et des schémas d'action flexibles – d'après (Edelman 1989:156-159).

Nous avons vu comment les différentes cartes interconnectées parvenaient à catégoriser les signaux d'entrée, externes ou internes : catégorisation extéro- ou intéroceptive, d'abord mono- puis multimodale. Les catégorisations des perceptions du monde extérieur n'acquièrent la pertinence que lorsqu'elles sont mises en relation avec l'état interne courant, i.e. avec la catégorisation des *valeurs* internes – cf. Figure II-3.

Une capacité essentielle est la catégorisation de la comparaison précédente (entre les C(W) et C(I)) dans des couches supérieures du cerveau, et la *mémorisation* des catégories dominantes obtenues, ou points fixes ; cela est réalisé dans certaines aires associatives du cortex (cf. aussi Figure II-4). Cette capacité est utilisée pour comparer les perceptions courantes C(W) avec les associations plus anciennes, déjà catégorisées et mémorisées **C[**C(W).C(I)**]**, à l'aide de multiples connexions réentrantes (en gras sur la Figure II-3). Cela conduit à une recatégorisation permanente de la mémoire des **C[**C(W).C(I)**]**, dont les unités *acquièrent ainsi le statut de concepts*.

Cette capacité de catégorisation des perceptions présentes à l'aide de catégories conceptuelles passées, fondées sur les valeurs internes, constitue la base de la *conscience primaire*, présente selon Edelman chez les vertébrés supérieurs. On comprend alors le titre de son ouvrage, *Le Présent Remémoré*, qui résumé précisément ce processus essentiel. La capacité conceptuelle permet de s'affranchir des comportements réflexes, génétiquement transmis, et moduler son comportement en fonction de l'expérience.

II.A.2.iv Les symboles, le langage et la conscience de niveau supérieur

Des étapes supplémentaires sont nécessaires pour rendre compte du langage. Le premier est la capacité à structurer la mémoire des concepts, et déceler des relations d'ordre ou de dominance ; il s'agit d'une « présyntaxe ». La *fonction symbolique* consiste à privilégier, de façon relativement arbitraire, un élément de la situation catégorisée et mémorisée (un « token » élément du concept), et à pouvoir établir un lien d'évocation entre cet élément et le concept mémorisé « tout entier ». La présence de tels éléments dans la mémoire de travail sera alors possible sans que la situation « symbolisée » soit réalisée dans l'environnement. La manipulation de ces *symboles* permet de s'affranchir du présent, de représenter et planifier des actions complexes, et de posséder une image de soi.

La capacité symbolique donne alors naissance à la *sémantique*, ou association entre symboles et concepts ; la capacité de traitement fonde la capacité syntaxique, et donc la fonction du *langage* tout entière.

> *Ainsi, la théorie repose sur l'hypothèse que pour construire une syntaxe riche, on doit déjà posséder les structures réentrantes qui fournissent la* base conceptuelle *de la sémantique. Selon la TSGN, cette base conceptuelle est fournie par la catégorisation de correspondances globales dans les régions frontale, temporale et pariétale, comme décrit [précédemment].* (Edelman 1989:176-177)

Ces capacités ne sont présentes que dans une seule espèce, la nôtre ; bien qu'Edelman ne s'intéresse pas directement à l'évolution naturelle, son exposé en décrit visiblement un déroulement possible, et nous verons au II.B.2.ii l'esquisse évolutionniste de Derek Bickerton(1990), compatible avec celle-ci. La description d'Edelman est étayée par de nombreux arguments neurologiques, qui restent consensuels – par exemple, en ce qui concerne les aires de Broca et de Wernicke, comparer la Figure II-4 avec la Figure II-1 du II.A.1.i. Une des principales différences est que les connexions suggérées par Edelman sont toutes réentrantes. La tripartition « concepts / syntaxe / sémantique-lexique » est donc une des propositions les plus précises pour des bases cérébrales du langage.

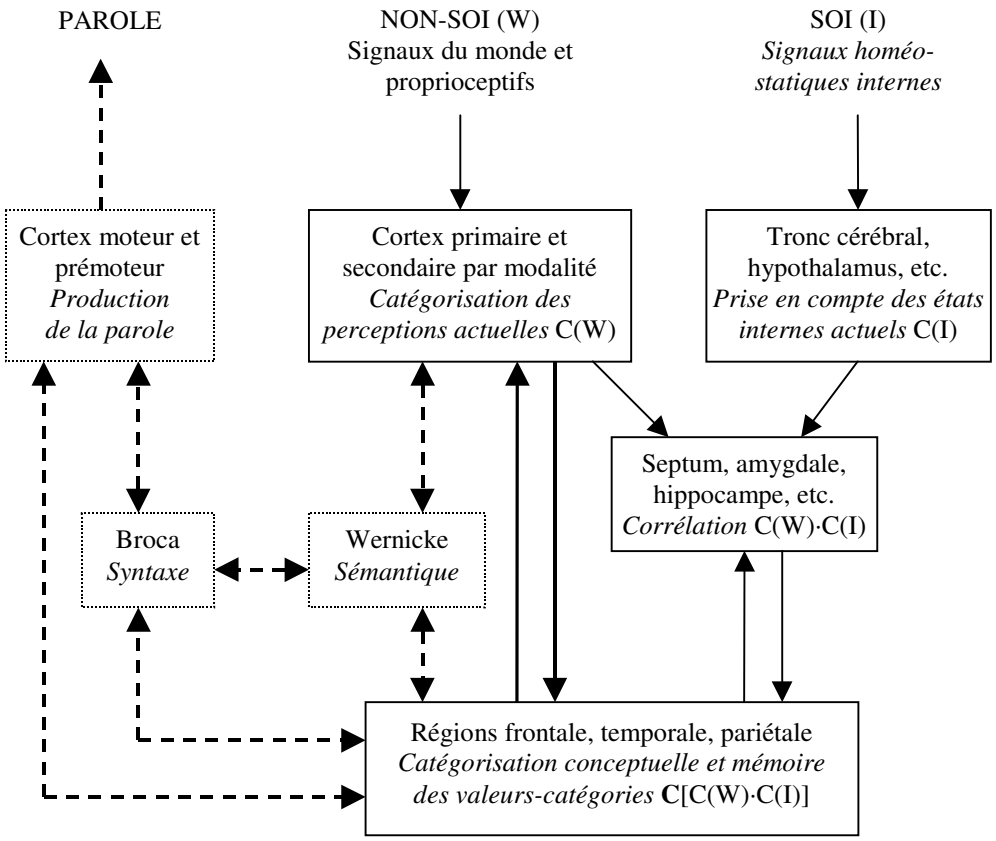


Figure II-4. Structures et fonctionnalités *nécessaires à la conscience primaire (boucle réentrante en* ***gras****), puis à la conscience de niveau supérieur et au langage (boucles réentrantes en pointillé). Synthèse d'après (Edelman 1989).*

II.A.2.v Une théorie prometteuse

Toutes les affirmations d'Edelman sont, certes, loin d'être justifiées, surtout lorsque l'on aborde les fonctions intellectuelles supérieures ; ici, la précision même des affirmations laisse parfois à désirer. On peut aussi, sur des points de détail, souhaiter une meilleure justification du « darwinisme » annoncé, avec une mise en évidence des éléments caractéristiques d'une théorie sélectionniste. Plus près de nos propres objectifs, on remarque aussi que cette théorie se focalise surtout sur les phénomènes perceptifs, et néglige quelque peu le fonctionnement moteur (aires motrices, cervelet, plans moteurs, syntaxe). Mais dans l'ensemble l'excellente intégration de la théorie ne peut que servir la modélisation, et les descriptions contiennent suffisamment d'éléments pour une modélisation de la fonction langagière dans les agents informatiques.

II.B L'émergence de l'interaction linguistique

Après avoir abordé les bases individuelles de la faculté de langage, nous passons maintenant à l'étude de son aspect *collectif*. Cet aspect ne fait pas encore l'objet d'un domaine de recherche en soi, bien que des sous-domaines s'y intéressent : la linguistique

historique, l'étude des dialectes et idiolectes, l'acquisition du langage, les modélisations informatiques (décrites au II.D.2 ci-dessous). Il est en effet difficile de caractériser ce qui, dans le langage, appartient à la collectivité des locuteurs, sans se réduire à l'individu. Afin de faire le point sur cette question essentielle, nous avons entrepris une étude synthétique des différentes situations où la faculté de langage émerge entre deux ou plusieurs individus.

II.B.1 Intérêt de l'étude des langages émergents

L'étude successive des cas d'émergence du langage apporte des idées essentielles pour les approches « émergentes » en intelligence artificielle, en particulier pour la modélisation multi-agent. Sur un plan méthodologique, cette étude permet d'abord de définir les frontières de la capacité linguistique elle-même : à partir de quel stade peut-on parler de « langage » ? En effet, où commence la « langue » sur le long chemin qui mène des états sans langue – dans toute leur diversité : animaux, bébés, aphasiques – à l'état avec langue – sujet humain normal ? Et, en même temps, y a-t-il des étapes intermédiaires bien identifiées, que l'informaticien pourrait viser comme sous-buts de ses simulations ? Nous verrons des réponses ci-dessous.

De façon plus appliquée, ces analyses éclaircissent aussi la question de savoir s'il est plus facile de modéliser directement l'état final (avec langue) ou bien l'état l'initial *et* le processus d'émergence, ou encore une quelconque combinaison proportionnelle de ces deux approches. Les tenants de l'IA symbolique défendent la première possibilité (Newell et Simon 1976). Toutefois, certaines approches alternatives ont apporté des solutions à des problèmes complexes en utilisant des algorithmes d'évolution artificielle, afin trouver des réponses ou configurations impossibles à trouver par d'autres moyens (exemples : réseaux de neurones, algorithmes génétiques). Ces mécanismes d'évolution peuvent être plus ou moins rapides et complexes – et l'« émergence » sera d'autant plus frappante que la méthode sera rapide et peu complexe.

L'étude de l'émergence du langage dans la nature permet d'estimer la complexité de l'état initial et celle du processus d'évolution dans plusieurs situations : apparition de la langue dans notre espèce (primates → hommes, II.B.2), acquisition d'une première langue par un individu (bébé → enfant parlant, II.B.3), créolisation (pidgin ou mélange linguistique → créole II.B.4). D'autres situations ne sont pas pertinentes ici : le changement linguistique (pas de progrès qualitatif), l'apparition de la communication animale (trop éloignée d'une langue, même pour les primates), l'acquisition d'une seconde langue (langue existante dans l'état initial, compétence finale discutable).

II.B.2 Évolution du langage dans l'espèce humaine

II.B.2.i Un domaine fort débattu

La question de l'origine du langage est fort ancienne : on peut penser au *Cratyle* de Platon, à la Tour de Babel biblique, etc.(Eco 1993). Depuis deux ou trois siècles cette

question s'est très fréquemment posée : il est d'usage de rappeler la position de la Société de Linguistique de Paris, au siècle dernier, consistant à refuser tout mémoire sur le sujet, les nombreuses spéculations se révélant en général invérifiables. Depuis une trentaine d'années toutefois, on assiste à un renouvellement des argumentations, à la lumière des découvertes en paléo-anthropologie, en neurosciences, en linguistique, et des nouvelles méthodes évolutionnistes (cf. aussi la question de l'organe du langage, II.A.1.iii).

Un problème assez ancien est de savoir si le langage est un produit de l'évolution biologique (comme la vision ou la station bipède) ou de l'évolution culturelle (comme la cuisson de la nourriture, ou les rites de mariage). Les nombreux éléments conventionnels du langage ont d'abord fait penser à l'évolution culturelle, mais des théories récentes rappellent l'importance du facteur biologique (Pinker 1994). Parmi leurs arguments on retiendra l'omniprésence du langage dans notre espèce, son ancrage cérébral (II.A.1), sa « redécouverte » par les enfants (acquisition, II.B.3, ou créolisation, II.B.4).

Une question qui suit est : si le langage est (aussi) le fruit d'une évolution biologique, quel est l'organe concerné ? On a soutenu que le langage avait été rendu possible par l'évolution de l'appareil phonatoire (mais cela semble restrictif, comme le montre la langue des signes), ou bien par le développement de la capacité crânienne ayant entraîné une intelligence accrue. Ces deux facteurs sont certes importants, surtout le second, l'interaction entre pensée et langage étant incontestable. Toutefois, il semble que le facteur essentiel soit l'évolution d'aires spécialisées pour la langage dans le cerveau, celles dites de Broca et Wernicke notamment.

Une question déjà plus technique est alors : quelle est la *fonction* qui a pu être renforcée par sélection naturelle pour aboutir au langage pleinement développé ? Deux points de vue extrêmes ont été défendus : le langage n'a aucune fonction adaptative, et a évolué comme sous-produit d'une autre adaptation (idées de S.J.Gould, défendues aussi par Chomsky) ; ou bien le langage a une fonction, qui reste à déterminer. On a proposé la fonction communicative (Pinker et Bloom 1990), ou un rôle dans la cohésion sociale des groupes, ou dans le contrôle moteur individuel, et surtout dans la *représentation interne* et le raisonnement. C'est ce dernier point que nous avons tendance à privilégier, puisqu'il est en accord avec les idées d'Edelman (II.A.2) et de Bickerton (II.B.2.ii ci-dessous).

Le renouvellement de la problématique des origines du langage est visible dans les conférences récentes tenues à ce sujet, par exemple (Hurford, Studdert-Kennedy et Knight 1998). Dans ces conférences, la modélisation informatique a une place de choix. Elle possède en fait un double statut : en tant qu'outil de modélisation, elle peut apporter des réponses expérimentales aux débats actuels, bien que ces validations demandent des précautions méthodologiques. Inversement, les modèles informatiques peuvent tirer parti des considérations évolutionnistes et utiliser des propriétés émergentes pour donner à des agents ou à des programmes des capacités langagières (cf. II.D.2).

En ce qui concerne cette deuxième possibilité, on pourrait objecter dès le départ qu'il semble impossible de modéliser en un temps raisonnable *l'ensemble* du processus évolutif, en partant des capacités que l'on sait implémenter explicitement (assez réduites) et en espérant faire évoluer des capacités supérieures (le langage). Toutefois, il n'est pas

certain qu'il soit plus simple de programmer directement le *résultat* de cette évolution. Les études de l'évolution du langage peuvent aussi aider l'informaticien à choisir des éléments réalistes pour une simulation : la fonction du code de communication, l'état initial et les états intermédiaires.

II.B.2.ii Théorie proposée par Derek Bickerton (1990)

Connu surtout pour ses travaux sur les langues créoles[9], Derek Bickerton est aussi l'auteur de deux livres plus généraux sur l'évolution du langage humain (Bickerton 1990, 1995) ; *Language and Species* est son livre le plus connu et le plus commenté[10]. Ce livre est d'un intérêt tout particulier pour ceux qui étudient la modélisation de phénomènes linguistiques à l'aide de populations d'agents, puisqu'il fournit, à titre d'hypothèse, quelques jalons précis dans l'évolution de la faculté de langage, et notamment un *état intermédiaire* particulièrement crédible, la *protolangue*.

II.B.2.ii.a Stades initial et intermédiaire de l'évolution

Selon Bickerton, le langage humain ne trouve pas son origine phylogénétique dans les systèmes de communication animale, mais dans les systèmes de représentation. En effet, on constate pour ces derniers une bonne continuité évolutive, alors que l'écart entre les cris d'alerte (ceux des singes, par exemple) et une langue semble incommensurable.

Bickerton analyse les formes qui pourraient ressembler aux stades précurseurs du langage humain, considérés souvent comme inaccessibles. Les singes peuvent apprendre un système limité d'associations signe/catégorie, mais jamais la capacité de structurer ces signes en un énoncé. Les stades à un mot/deux mots dans l'acquisition d'une première langue par l'enfant présentent des similitudes frappantes avec la situation des singes, comme le montre une comparaison de dialogues (Bickerton 1990:110). Il y a aussi le cas d'enfants séquestrés, qui peuvent acquérir tardivement un lexique assez riche, mais pas les moyens de structurer leurs énoncés (items grammaticaux, ordre). Enfin, les pidgins, langues émergentes de contact qui ne sont la première langue d'aucune population, partagent certaines caractéristiques avec les autres exemples.

Ces quatre cas conduisent à l'idée qu'il existe un mode de communication langagière plus rudimentaire que le langage standard. Ce mode, sous-jacent au langage standard et cérébralement plus robuste, redevient apparent dans certains cas d'aphasie. Bickerton appelle ce mode *protolangage*, et donne cinq propriétés qui lui manquent par rapport au plein langage : la signifiance de l'ordre des mots, l'existence d'éléments inexprimés (pronoms nuls par exemple), le nombre fixe d'arguments pour chaque verbe, la construction récursive des syntagmes ou phrases, l'omniprésence des items grammaticaux (morphologiques ou syntaxiques).

---

[9] Sur lesquels nous reviendrons ci-dessous au II.B.4, et à propos de la référence, au **VI.A.2**.

[10] En français *Langage et Espèce* (Bickerton 1990), ouvrage dont nous avons publié un compte-rendu (Popescu-Belis 1998).

II.B.2.ii.b Esquisse de l'évolution

Le protolangage serait donc le principal état intermédiaire dans l'évolution du langage. Bickerton tente de relier les données paléontologiques à son modèle d'évolution du langage en deux étapes : (1) lexicalisation d'un système de représentation déjà présent chez les singes pour arriver au protolangage ; et (2) apparition des propriétés syntaxiques conduisant au langage.

Ainsi, à l'*Homo habilis* (-2.5 à -1.6 Ma) dépourvu de langage succède l'*Homo erectus* (-1.6 Ma à -290.000 ans ou moins) muni de protolangage. Pour la transition vers la syntaxe et l'apparition de l'*Homo sapiens*, Bickerton défend un *scénario catastrophique*, qui attribue cette apparition à une seule mutation survenue en Afrique (le scénario *catastrophique* et le scénario *graduel* font l'objet de vifs débats en paléontologie). Bickerton suggère que la mutation a entraîné plutôt un mode nouveau de connexion neuronale et non l'apparition d'une aire nouvelle dans le cerveau. L'*Homo erectus* pouvait probablement produire rapidement de nombreux mots se rapportant à une situation, mais dans le désordre. La mutation a peut-être réalisé la conjonction de deux capacités permettant d'organiser la « salade de mots » :

- la structure de syntagme nominal, apparue afin d'ordonner les trois éléments nécessaires pour faire référence à une entité : sa classe, ses propriétés particulières et sa spécification en nombre, familiarité, proximité, etc.
- la structuration de toute situation perçue en arguments autour d'un verbe unique.

L'utilisation de la structure de syntagme pour décrire l'ensemble d'une situation et insérer de façon canonique et récursive le verbe muni de ses arguments dans l'énoncé serait donc le résultat de cette unique mutation.

II.B.2.ii.c Comparaison avec d'autres travaux

La théorie de Bickerton reste pour l'heure assez hypothétique. Toutefois, on remarque un certain accord avec les idées d'Edelman sur la prééminence de la fonction conceptuelle/sémantique sur la fonction syntaxique. Les deux esquisses sont complémentaires, Edelman étant plus précis sur l'aspect neurologique, et Bickerton sur l'aspect linguistique et évolutif.

« MODE PRAGMATIQUE »	« MODE SYNTAXIQUE »
• structure thème/rhème	• structure sujet/prédicat
• conjonction floue des termes	• subordination stricte des termes
• taux de transmission faible	• taux de transmission élevé
• ordre des mots « pragmatique » : la nouvelle information suit l'ancienne	• ordre des mots signalant les cas sémantiques
• autant de noms que de verbes, verbes sémantiquement simples	• plus de noms que de verbes, verbes sémantiquement complexes
• pas de morphologie grammaticale	• morphologie grammaticale élaborée
• l'intonation/accentuation marquent l'information nouvelle (focalisation)	• intonation/accentuation faibles, parfois absentes

Table II-1. Deux modes de langage selon Givón (1979:98).

Il est utile aussi de comparer le couple protolangue/langue aux deux pôles d'utilisation d'une langue, pragmatique/syntaxique, mis en évidence par Talmy Givón (1979:98 sq.), cf. Table II-1. Dans toute langue, on constaterait diachroniquement une propension vers le mode syntaxique ; la même évolution aurait eu lieu pour la faculté de langage, et se retrouverait aussi dans le passage des pidgins aux créoles. Givón conçoit donc le mode pragmatique comme antérieur au mode syntaxique, mais ses hypothèses sont plus spéculatives et moins précises que celles de Bickerton[11].

Bickerton a donc le mérite d'avoir isolé un nombre réduit de principes qui font la spécificité de la syntaxe, et d'avoir réduit son apparition à un seul pas, dû à une mutation heureuse; le scénario graduel, à l'inverse, a du mal à expliquer comment on peut développer d'abord une fraction de la capacité syntaxique. Pour ceux qui étudient la modélisation de cette émergence, le stade du *protolangage* et la *mutation syntaxique* sont des éléments originaux et exploitables.

II.B.3 Acquisition du langage par l'enfant

L'apparition des compétences linguistiques chez l'enfant doit être une autre source d'inspiration privilégiée pour la modélisation informatique. Comme pour l'évolution du langage, on peut distinguer une modélisation visant à évaluer les théories de l'acquisition existantes, et une modélisation visant à simuler le processus afin de fournir de façon aisée et réaliste des connaissances linguistiques à un programme. Sans pouvoir entrer dans le détail des nombreux travaux réalisés (surtout si on leur ajoute tous les programmes de TAL capables d'apprentissage) nous allons montrer comment les théories actuelles de l'acquisition peuvent orienter la modélisation informatique.

[11] Les deux chercheurs ont collaboré au moins une fois. Nous verrons les travaux de Givón sur la référence (**V.C.1**), et un schéma développé conjointement avec Bickerton (**VI.B.2**).

II.B.3.i Données psycholinguistiques

L'acquisition du langage a été par excellence un problème psychologique, tant ses manifestations sont sujet d'observation, d'analyse et d'interprétation. L'ancienneté du problème – on rappelera l'expérience du pharaon Psammétique, rapportée par Hérodote (*Histoires*, II, 1) – fait que de nombreuses études psycholinguistiques existent déjà ; toutefois, l'interprétation des données est encore sujette à discussions, et leur explication reste à venir, car elle dépasse le cadre de la psychologie et dépend aussi d'une théorie physicaliste du langage.

II.B.3.i.a Chronologie

L'acquisition du langage par l'enfant suit une progression étonnamment uniforme selon les individus, les cultures et les langues respectives (Slobin 1984). Celle-ci semble dépendre assez peu des efforts que les parents font pour parler aux enfants (Lieven 1994), ou du nombre de langues apprises (enfants bilingues). Nous synthétisons dans la table ci-dessous les principales étapes du processus, d'après (Kandel, Schwartz et Jessell 1995) – tout en faisant remarquer le caractère artificiel de la séparation des capacités phonétiques, lexicales et syntaxiques, et la différence importante entre les capacités de perception/compréhension et de production.

Âge	Capacités phonétiques	Capacités lexicales	Capacités syntaxiques
6-15 sem.	début du *babil*		
3-8 mois	*babil* riche		
1 an	le *babil* s'estompe ; quelques exclamations	4–5 *fonctions* pour les exclamations	
1an 1/2	pauvreté (contrastant avec le *babil*)	30 à 50 mots : noms, adjectifs, verbes d'action	*holophrases* (phrases à un mot)
2ans	lente amélioration ; état provisoire	50 à quelques centaines	*style télégraphique* (phrases à deux mots)
2ans 1/2	idem	700 à 800 mots (proportion de noms quatre fois supérieure à celle de l'adulte)	phrases à trois mots et plus ; nombreuses fautes
3 ans	presque adulte	un millier et plus	phrases bien formées
4 ans	quasi adulte	proche de l'adulte ~3000 mots à 4 ans (adulte : ~10000)	proche de l'adulte

Table II-2. Étapes dans l'acquisition du langage, d'après (Kandel, Schwartz et Jessell 1995).

Deux indices globaux de performance servent à mesurer le niveau atteint. On peut estimer le *volume du lexique acquis*, en général pour la production[12]. Ce volume a une croissance quasi exponentielle entre 1an1/2 et 4 ans (moyenne de 3 mots nouveaux par jour). Plus difficile à mesurer, *la longueur moyenne d'une énonciation* croît en général par paliers, jusqu'à l'âge de cinq ans ou plus ; les premiers paliers sont à l'origine des stades dits « à un mot » et « à deux mots ». Par ailleurs, ces mêmes étapes se rencontrent aussi chez les enfants sourds-muets, avec une langue des signes.

Le mode sur lequel les parents s'adressent aux enfants a fait l'objet de nombreuses études (Gallaway et Richards 1994), et est appelé « mamanais » (en anglais « motherese »). Il s'agit d'un langage aux contours sonores très accentués, et d'un niveau lexical et syntaxique relativement simple. Bien qu'il soit relativement répandu dans le monde occidental, ce mode ne semble ni universel, ni indispensable à l'acquisition (Lieven 1994).

Les « premiers mots » de l'enfant posent déjà un problème de classification, puisqu'ils servent souvent de substitut à une expression correcte plus complexe, dont l'enfant n'est pas encore capable ; par exemple, on rencontre souvent des « fonctions » exprimées par « y en a plus », « encore », « au revoir ». D'autre part, même pour les noms communs, leur production (émission) revêt souvent de façon implicite une attitude plus complexe envers l'objet désigné (désir, crainte, commentaire) ; nous reviendrons sur les noms et la désignation au VI.A.1.ii. On peut retenir qu'à ce stade trois classes d'expressions sont présentes : *noms d'objets, expressions de sentiments et de besoins*.

II.B.3.i.b Un état intermédiaire : le style télégraphique

Le stade « à deux mots » est un état intermédiaire remarquable. L'analyse de nombreux enregistrements et transcriptions montre en effet que l'on peut isoler une période de l'acquisition où la longueur moyenne des énonciations est voisine de deux (Brown 1973, Demopoulos et Marras 1986). Ces énonciations contiennent à la fois des mots de la classe ouverte (CO) et des mots de la classe fermée, ou pivot (PV)[13]. Les schémas prédominants semblent être CO+CO, et un peu plus tard PV+CO ; des mots de la classe ouverte sont acquis en premier, puis certains PV, malgré leurs proportions très différentes dans les exemples entendus. Il semble aussi qu'il y ait déjà une forte préférence à reproduire la syntaxe de la langue apprise dans l'ordre des deux mots (Pinker 1984, Slobin 1984), par exemple sujet/verbe/objet pour l'anglais ou le français[14].

---

[12] C'est-à-dire le vocabulaire « actif », émis par l'enfant, par rapport au vocabulaire « passif » ou compris. Il y a toutefois de bonnes raisons de penser que les capacités de compréhension sont en avance sur les capacités de génération.

[13] Cette distinction assez répandue sépare les mots grammaticaux, ou foncteurs, des autres mots, dits « pleins ». Les premiers sont en général courts, fréquents, et en nombre très limité pour une langue donnée (prépositions, auxiliaires, terminaisons fonctionnelles, etc.). Les seconds sont les noms, les adjectifs, les verbes, auxquels la langue ajoute aisément, en permanence, de nouveaux membres (comme « chromosome », « ordinateur » ou « mél »).

[14] La prédominance de la forme PV+CO s'explique par une tendance à l'ergativité : le sujet d'un verbe intransitif et l'objet d'un verbe transitif ont un statut et une importance semblables. L'ergativité, qui est une

II.B.3.i.c Quatre situations exceptionnelles

Ajoutons ici quatre situations d'acquisition exceptionnelles :

- la créolisation – des enfants exposés à un code de communication imparfait (pidgin) développent une véritable langue, cf. II.B.4 ci-dessous ;
- les enfants sourds de parents entendants – non exposés à une langue des signes, ils parviennent à en développer une pour communiquer avec leur mère, cf. II.B.3.iii ;
- langage privé des jumeaux (Pinker 1994) – on a observé des rares cas où des jumeaux développent un langage qui leur est propre (essentiellement à l'aide de conventions lexicales) ;
- enfants isolés – non exposés à une interaction linguistique (« enfants-loups », enfants séquestrés, etc.), les enfants ne développent pas de langue. L'apprentissage ultérieur est relativement efficace pour le lexique, mais s'avère impossible pour la syntaxe ; les sujets atteignent un « protolangage », tel celui qui est supposé par Bickerton (II.B.2.ii.a).

Ces exemples montrent que même en présence d'informations linguistiques extrêmement réduites, les enfants sont capables de développer un langage, ou bien qu'ils peuvent en développer un qui soit assez éloigné de ce qu'ils entendent. Le dernier exemple montre qu'un déclencheur d'interaction linguistique *doit* être présent pour que cela se produise, ce qui ne correspond pas aux modélisations informatiques actuelles (II.D.2).

II.B.3.ii Expliquer l'acquisition

II.B.3.ii.a Les arguments

On retrouve dans le débat sur l'acquisition la ligne de fracture esquissée au II.A.1.iii ci-dessus : inné ou appris, faculté spécifique ou manifestation de l'intelligence. Parmi les arguments en faveur du « caractère appris » on cite : chaque enfant apprend la ou les langues de son milieu ; l'enfant acquiert bien certains universaux cognitifs en dehors du langage (par exemple la causalité, les nombres) ; on voit mal comment le langage serait codé génétiquement ; de nombreux éléments lexicaux et syntaxiques semblent bien objet de convention.

En faveur de bases biologiques spécifiques (« caractère inné ») on cite : la spécificité du langage à l'espèce humaine ; l'existence de nombreux invariants linguistiques, que ce soit en phonétique, morphologie, syntaxe, à la lumière desquels la variabilité du langage apparaît plus restreinte ; la similarité du processus d'acquisition chez tous les enfants (même vitesse, mêmes étapes), y compris sourds-muets (avec une

---

propriété de certaines langues, a été remarquée aussi dans le code de communication gestuel spontané entre enfants sourds-muets (cf. (Goldin-Meadow et Mylander 1998) et notre II.B.3.iii). Il semble toutefois que, pour les langues où les éléments PV sont syllabiques et accentués, leur première utilisation par l'enfant est plus précoce (Slobin 1984).

langue des signes) ; l'existence d'une phase critique, hors de laquelle l'acquisition est peu efficace. Un des arguments les plus fréquemment cités est *la pauvreté du stimulus* : la complexité du langage « acquis » dépasserait largement ce que les exemples entendus par l'enfant ont pu lui enseigner.

Encore une fois, cette dichotomie est en soi assez artificielle. Des aspects de chaque point de vue sont sans doute proches de la vérité, mais à l'heure actuelle le côté « inné » semble prendre de l'importance, et le côté « appris » semble reculer – si cette comparaison a un sens. En somme, très peu de choses seraient nécessaires pour déclencher le développement du langage, mais il y a un minimum d'éléments que l'on ne peut supprimer[15].

II.B.3.ii.b Les théories

Il n'est pas dans notre propos de rendre compte de tous les modèles théoriques proposés pour l'acquisition : nous nous limiterons à quelques exemples importants.

Les théories formelles de l'apprentissage développées en intelligence artificielle ont aussi été appliquées au langage. Ainsi, les travaux de Gold (1967) ont été à la base des recherches en « apprenabilité » ; ces travaux définissent en termes formels l'apprentissage d'un langage (chaînes de caractères autorisées ou non) et évaluent le nombre d'exemples nécessaires. Un des premiers résultats est que seulement des langages finis peuvent être appris sans rétroaction négative (cas des enfants recevant peu de corrections). Des objections sur l'application de ce formalisme aux situations réelles ont été soulevées, par exemple par (Pinker 1984). Celui-ci a toutefois proposé d'autres formalismes pour exprimer le problème de l'acquisition, et notamment les concepts réciproques de *conservatisme* et de *productivité* (Pinker 1986), ainsi que la notion de *surgénéralisation* ; le formalisme semble seulement descriptif, sans préjuger d'un mécanisme interne.

Le point de vue chomskyen (II.A.1.iv ci-dessus) apporte une proposition élégante. En effet, la théorie des principes et paramètres (P&P) affirme que les principes de la « grammaire universelle » sont innés, ainsi que le type des paramètres. Par exemple, parmi les principes pourrait figurer une syntaxe X-barre (Jackendoff 1977), un paramètre étant la position (droite/gauche) de la « tête » (branche de l'arbre X-barre qui s'oppose aux modifieurs), ou le caractère obligatoire ou non du sujet. L'acquisition consisterait alors, sur la base des exemples entendus, à donner aux paramètres les valeurs correspondant à la langue apprise, et acquérir le lexique. Plusieurs exemples de principes et paramètres ont été proposés, mais le principal problème semble être la façon dont les principes pourraient être innés : peu de propositions existent.

Nous citerons aussi une hypothèse quelque peu analogue, inspirée à Derek Bickerton par ses études sur les langues créoles, et que celui-ci appelle « l'hypothèse du bioprogramme » (Bickerton 1984). Très schématiquement, par rapport à la théorie P&P, Bickerton suggère que les capacités grammaticales innées seraient munies de valeurs par

[15] Les enfants grandissant seulement devant un poste de télévision n'acquièrent pas le langage (Pinker 1994:278)

défaut, qui se manifestent lorsque les enfants « inventent » une langue créole (cf. II.B.4 ci-dessous) ; certains traits sémantiques figurent également dans ce répertoire, comme par exemple une distinction des groupes nominaux en spécifique / non-spécifique que nous examinerons plus en détail au VI.A.2. Il y aurait donc une « langue par défaut » qui serait inhibée par les contraintes de la langue apprise, témoin certaines erreurs systématiques que commettent les enfants apprenant une langue non créole. Ces idées, très débattues, sont un autre prolongement des théories créolistes de Bickerton, ayant précédé ses idées plus générales sur l'évolution du langage.

II.B.3.iii Un cas exceptionnel : les enfants sourds de parents entendants

L'étude des cas où le stimulus proprement linguistique perçu par l'enfant est inexistant permet d'estimer d'autant mieux la contribution de l'enfant au développement du langage, ou bien le rôle des contraintes universelles portant sur les systèmes de communication.

Les enfants ayant un fort déficit auditif sont souvent diagnostiqués assez tard, et certains parents ne peuvent ou ne désirent pas leur enseigner une langue des signes, qui leur permettrait pourtant de développer pleinement leur potentiel grammatical[16]. Le cas de ces enfants a été étudié notamment par Susan Goldin-Meadow et Carolyn Mylander (Goldin-Meadow, Butcher, Mylander et Dodge 1994, Goldin-Meadow et Mylander 1990, Goldin-Meadow et Mylander 1998, Goldin-Meadow, Mylander et Butcher 1995). L'analyse des interactions pseudo-langagières enregistrées, qui dépasse notre propos, fait toute la difficulté des expériences citées.

Pour les besoins quotidiens, ces enfants développent, en général de leur propre initiative, un code de communication par signes avec leur mère. Ces signes s'apparentent aux noms, verbes, et adjectifs naturels, auxquels s'ajoutent des pointages et des signes « grammaticaux » ; leur structuration, tant au niveau « morphologique » qu'au niveau « syntaxique » montre certaines régularités dans les cas étudiés, et cela même à travers plusieurs langues de base des parents (Goldin-Meadow et Mylander 1998). Ces régularités grammaticales dépassent largement celles présentes dans les gestes produits par la mère.

Les auteurs synthétisent dans (Goldin-Meadow et Mylander 1990) un certain nombre d'éléments qui seraient apportés par l'enfant à son code de communication gestuel :

- les assertions sont structurées en gestes prédicatifs accompagnés d'un nombre fixe d'arguments gestuels, dans un ordre précis. La propension des enfants à organiser leurs assertions sur un mode prédicatif (prédicat + arguments) a été aussi constatée pour les enfants normaux, et peut refléter une certaine organisation du système perceptif/conceptuel ;

[16] Pour répondre aux éventuelles interrogations du lecteur, les parents qui choisissent de ne pas exposer leurs enfants à une langue signée espèrent souvent leur enseigner une langue phonétique (plus propice à l'intégration sociale) grâce à la lecture sur les lèvres et à l'orthophonie.

- la focalisation, dans une assertion, porte sur le *patient* du prédicat, à savoir l'objet d'un prédicat transitif. Pour les prédicats intransitifs, *l'agent* est traité d'une manière semblable au *patient* du prédicat transitif. En outre, les gestes pour les *patients* sont plus fréquents que ceux pour les *agents*, et attirent spatialement vers eux les autres gestes ;

- les rôles de *patient* et *agent* sont distingués aussi dans la morphologie des signes : ceux-ci sont infléchis selon leur rôle. Une certaine forme d'accord grammatical entre signes a aussi été constatée ;

- enfin, la récursivité a été observée, c'est-à-dire l'emboîtement de séquences de signes dans des structures plus complexes que celles produites par les parents.

Naturellement, une exploitation plus détaillée de ces résultats est nécessaire pour dégager les propriétés d'un « code de communication spontané », ou « langage minimal » qui pourrait être simulé dans une société d'agents. Une question qui se pose est certainement l'origine des capacités observées : une structuration universelle du système cognitif ou représentationnel peut expliquer certains éléments, mais d'autres semblent reliés à une fonction linguistique ou communicative spécifique.

II.B.3.iv Conséquences pour la modélisation par agents

L'ensemble de ces développements nous permettent de mettre en avant les problèmes essentiels que les modélisations multi-agent de l'acquisition seront amenées à résoudre. D'abord, comme pour l'évolution, on peut encore une fois soutenir qu'il est plus simple de programmer directement l'état final, et non son émergence.

Si on s'intéresse à celle-ci, un facteur essentiel pour l'informaticien est la proportion d'inné et d'acquis. Si l'enfant était seulement capable d'un apprentissage générique intelligent, et que le langage résidait dans les conventions de la société, alors peut-être qu'un programme d'apprentissage générique pourrait acquérir le langage – telle est par exemple la motivation des premiers travaux sur l'apprenabilité formelle.

Il semble toutefois que les structures cérébrales de l'enfant contraignent plus sérieusement le langage acquis, comme dans la théorie P&P par exemple. La question est alors : peut-on donner une modélisation réaliste, ou un équivalent computationnel, du cerveau ou esprit de l'enfant lorsqu'il commence à acquérir le langage ? Encore une fois, il y aura ceux qui tenteront de réaliser cela directement, et ceux qui remonteront encore le temps, pour partir d'un état encore plus précoce (l'œuf fécondé ?) et faire évoluer un cerveau ou esprit capable d'apprendre. Tout cela relève de la spéculation, mais ce que l'étude de l'acquisition montre, c'est que *seule une structure complexe et contraignante permet d'acquérir le langage en un temps raisonnable*.

Ces études démontrent aussi l'importance de l'interaction langagière dans l'acquisition et donc confortent l'idée qu'*il faut simuler un agent linguistique situé*, capable d'*utiliser la communication pour satisfaire ses besoins*. La section suivante

montre, sur l'exemple de la créolisation, que l'instruction d'un tel agent peut se faire à l'aide d'un agent plus rudimentaire.

Enfin, ces études décrivent de façon précise des étapes intermédiaires, qui peuvent servir d'objectifs pour l'informaticien. On ne peut prouver que ces points de passage sont nécessaires, mais il serait intéressant d'évaluer les programmes sur une telle échelle (Table II-2). Quel âge donnerait-on aux programmes actuels ? Peut-on construire un test de Turing de l'acquisition ? Ces questions montrent encore une fois que seuls des agents situés, interagissant entre eux et avec l'utilisateur à travers un environnement commun, peuvent prétendre à un classement sur une telle échelle.

II.B.4 Naissance d'une langue : la créolisation

Nous arrivons enfin au troisième cas d'émergence, le plus surprenant. L'étude des langues créoles a été renouvelée et popularisée de façon considérable par les travaux de Derek Bickerton (1975, 1981). Avant de montrer en quoi ceux-ci intéressent les créolistes[17], les linguistes et les informaticiens, il est nécessaire de présenter brièvement la problématique.

II.B.4.i Généralités sur les langues créoles

Le terme de langues « créoles » est appliqué aux langues apparues dans les anciennes colonies (surtout européennes) comme résultat d'un brassage important de populations d'origines diverses : principalement des descendants des anciens esclaves, puis ceux des colons, et éventuellement des indigènes. L'application même du terme à plusieurs langues souvent indépendantes montre bien le fait marquant du domaine : les nombreuses ressemblances de ces langues. Un des problèmes clés de la créolistique est d'expliquer les similarités observées, autrement dit de remplacer la définition extensionnelle des *créoles* (l'énumération des langues créoles) par une définition intensionnelle, fondée sur les propriétés caractéristiques de ces langues. Rappelons l'existence des *pidgins*, qui sont des langues de contact quasi dépourvues de grammaire.

Un grand nombre d'observations et de comparaisons sont nécessaires, c'est pourquoi ce domaine est très technique. Les propriétés communes concernent en général la grammaire des langues créoles, par exemple le système des articles ou des temps, modes et aspects verbaux. Le lexique est bien plus variable, puisqu'il est en général emprunté à une langue *base*, souvent la langue dominante des colons. À côté de la *base*, on distingue le *substrat*, qui est la langue ou famille parlée majoritairement par les nouveaux arrivants.

[17] L'étude des langues créoles est un domaine très riche, sujet à de nombreuses controverses. Une encyclopédie récente contient ainsi deux articles consécutifs : « Pidgins and Creoles » et « Pidgins, Creoles and Change », dans (Asher et Simpson 1994:3177-81 et 3181-86). Parmi les ouvrages détaillés, on peut citer ceux, en France, d'Albert Valdman (1978, 1981) et de Robert Chaudenson (1979), ainsi que ceux de Pieter Muysken (Muysken et Smith 1986).

On oppose en général deux explications des points communs observés – sachant d'emblée que la *base* ne peut en être l'origine, car on rencontre des créoles à base française, portugaise, anglaise, hollandaise, arabe, etc. La théorie des *substrats* soutient que les traits communs sont ceux des grandes familles linguistiques des nouveaux arrivants, en général esclaves africains. La théorie *universaliste* soutient que ces traits trahissent des tendances innées, universelles chez les humains, ou du moins des tendances universelles de la communication langagière.

II.B.4.ii Un scénario de créolisation

Derek Bickerton a proposé une définition intensionnelle du « créole », fondée sur sa description et son explication du phénomène de créolisation. La définition semble mieux acceptée que l'explication – faisant de Bickerton un rénovateur du domaine – et nous allons la résumer ici.

Lorsque des populations parlant des langues (très) différentes viennent à se rencontrer de façon nécessaire et prolongée, les adultes peuvent créer, pour les besoins de la communication, un code de communication fondé le plus souvent sur des emprunts lexicaux communs. Il s'agit d'un *pidgin* (ou *sabir*), sorte de « code » à morphologie et syntaxe floues ou inexistantes. Lorsque des enfants grandissent dans un milieu parlant pidgin, on remarque une tendance à la régularisation de ce code, et le développement de propriétés grammaticales ; ce processus peut prendre plusieurs générations si les langues originelles continuent à être parlées dans les populations.

Selon Bickerton, ce processus peut aussi avoir lieu en une seule génération, comme au début de notre siècle à Hawaii (son exemple de choix). Pour cela, il faut imaginer des enfants élevés soit par des couples linguistiquement mélangés, soit dans une collectivité mixte, et soumis majoritairement au pidgin. Le code de communication qui se forme dans le groupe des enfants possède des propriétés de régularité qui en font une véritable langue, et qui semblent partagées par de nombreuses langues créoles. Il faut bien sûr noter que les items lexicaux sont empruntés au pidgin (qui les a lui-même empruntés à la langue base), certains étant détournés pour servir de foncteurs grammaticaux.

II.B.4.iii Importance pour la modélisation informatique

La créolisation peut constituer une source d'inspiration essentielle pour la modélisation par agents de l'émergence du langage. En effet, elle montre que, à un stade précis de maturation, les « agents linguistiques naturels » (les enfants) peuvent utiliser un lexique existant pour développer une langue. Dans la nature donc, la détermination structurale des agents est telle qu'un code de communication *régulier* ou grammatical peut se développer spontanément – pourvu qu'un lexique existe déjà, et soit utilisé dans des *situations pertinentes*. La pertinence des interactions et leur médiation linguistique sont essentielles, mais non la « grammaticalité » de ce code. Il semble en revanche que la structure des agents qui apprennent doive être suffisamment riche et contrainte pour

exprimer des principes grammaticaux. Nous verrons une implémentation possible de ce paradigme d'émergence du langage au Chapitre III.

* * *

Après ce passage en revue des trois cas où le langage émerge dans la nature, le moment est venu d'examiner comment les programmes informatiques ont essayé de reproduire les différentes fonctions résultant de l'émergence, et plus récemment le processus lui-même. Encore une fois, nous séparerons les études de programmes ou agents isolés des réalisations concernant les sociétés d'agents communicants.

* * *

II.C Modélisations d'individus isolés

La notion d'isolement semble contredire l'idée même d'agent communicant. Toutefois, les stades précurseurs de la modélisation multi-agent se limitent parfois à l'étude de fonctions essentielles à l'autonomie de ces agents, sans aborder la communication. Nous examinons ici quelques exemples d'agents autonomes ou robots ayant été conçus dans la perspective des modèles multi-agent. À ceux-là s'ajoutent certains programmes de traitement de la langue qui se présentent aussi comme des « agents », même si leur autonomie est insuffisante.

Nous partirons donc des modèles d'agents autonomes et situés – donc convenant à une modélisation réaliste d'agents linguistiques – mais en fait dépourvus de capacités cognitives élaborées (II.C.1). À l'autre extrême, nous verrons que les programmes faisant montre de certaines capacités langagières sont peu autonomes, ou ne sont pas situés dans un environnement (II.C.3). Entre les deux, en attendant le développement d'agents linguistiques réalistes, nous verrons les modèles réalisés dans l'équipe de Gerald Edelman (II.C.2).

II.C.1 La notion d'animat

II.C.1.i L'autonomie en robotique

La robotique a été le champ de nombreuses réalisations d'entités situées, en interaction avec l'environnement. Une des priorités actuelles est *l'autonomie des robots*, en retard sur des capacités mieux maîtrisées, comme les organes moteurs (travaux en rapport avec la mécanique) ou les organes sensoriels (travaux en vision artificielle par exemple). Des techniques d'intelligence artificielle peuvent être employées pour le contrôle des robots (mécanismes de décision, planification), mais avec certaines limitations, à savoir le manque de flexibilité, la faible pertinence des réponses aux situations nouvelles, le travail important de modélisation préalable des connaissances – bref, une autonomie réduite du système.

Pour donner une idée des réalisations dans ce domaine, on peut mentionner la *RoboCup*, compétition de football entre robots issus de laboratoires de recherche. Des problèmes variés sont ici à résoudre, selon la classe choisie : simulation, petits robots commandés par un ordinateur, robots moyens totalement autonomes. Les meilleures équipes étudient déjà la coordination entre joueurs, mais des problèmes d'apparence triviale comme la reconnaissance des buts, de la balle ou le positionnement subsistent encore. En marge de ce véritable paradigme de recherche, le robot quadrupède élaboré par *Sony*, outre sa démarche réaliste, parvient à pousser une balle dans les buts. Rappelons enfin que la mission *Pathfinder* sur Mars ne pouvait être téléguidée en temps réel à cause des délais (une dizaine de minutes environ), nécessitant donc une autonomie à court terme du robot explorateur.

II.C.1.ii Simulation d'entités vivantes ou animats

Pour accroître l'autonomie des robots, on a étudié l'imitation d'être vivants adaptatifs simples, placés dans un environnement simulé informatiquement. Certes, les être vivants imités, par exemple les insectes, sont le fruit d'une évolution aussi longue que la nôtre et leur simplicité est toute relative. La biologie décrit leurs structures de contrôle, formées en général de ganglions nerveux déterminés génétiquement. Bien que simples, ceux-ci permettent à un insecte de se déplacer avec bien plus d'agilité que les robots actuels, et de déclencher une gamme de comportements appropriés afin d'assurer sa survie, par exemple la recherche de nourriture – un exemple d'*autonomie* que l'on cherche à reproduire.

Le terme *animat*, « animal + robot », a été forgé par Jean-Arcady Meyer (ENS, puis LIP6, Paris) pour désigner les agents ainsi réalisés. Un des travaux exemplaires de son équipe consiste à faire évoluer une structure neuronale de contrôle d'un animat hexapode capable de se déplacer, qui doit se diriger vers une source de nourriture (Kodjabachian 1998, Kodjabachian et Meyer 1998). Les premières générations ne possèdent qu'une cellule dans leur structure de contrôle ; cette structure évolue peu à peu avec les générations successives pour contrôler la marche hexapode et diriger l'animat vers la ressource. Le mode d'évolution de ces réseaux de neurones est remarquable : ce ne sont pas uniquement les connexions et leurs poids qui font l'objet d'un algorithme génétique, mais le *programme de construction* du réseau, construction ayant lieu dans la *phase de développement* de l'animat[18].

Des structures plus élaborées ont été proposées pour la vision d'un animat par l'équipe de Philippe Gaussier (ENSEA, Cergy), pour des animats simulés puis pour des robots réels (Gaussier, Revel, Joulain et Zrehen 1997, Gaussier et Zrehen 1995, Joulain, Gaussier et Revel 1996). L'accent est mis ici sur l'apprentissage associatif entre des configurations du champ visuel (« indications ») et des cibles à atteindre dans l'environnement. Un algorithme est proposé pour réaliser cet apprentissage dans le réseau de neurones.

[18] Une revue des approches « neurogénétiques » se trouve dans (Kodjabachian 1995).

II.C.1.iii Expérimentation avec des microrobots

Les chercheurs ont naturellement pensé à expérimenter en grandeur nature les structures de contrôle proposées pour les animats. Les robots *Khépéra* (EPF Lausanne) ont été conçus pour offrir un outil expérimental abordable (Mondada, Franzi et Ienne 1993), étant donné qu'aucune simulation ne peut reproduire la *complexité* d'un environnement naturel, fut-il simple (table + murs), et les *contraintes* que celui-ci impose. Il s'agit de petits robots cylindriques (d=6 × h=3 cm) munis de capteurs infrarouge, d'une microcaméra et de moteurs, disponibles commercialement. Une architecture fondée sur des réseaux récurrents servant à contrôler les déplacements de ces robots a été expérimentée par les concepteurs (Floreano et Mondada 1996).

Le point important est l'accent qui est mis sur la structure de contrôle, et non sur d'éventuels capteurs ou effecteurs sophistiqués. L'architecture *Khépéra* est tout entière orientée vers l'étude du contrôle, les problèmes mécaniques ou optiques étant relégués au second plan. Enfin, de l'aveu des auteurs, la simplicité et la taille du robot permettent d'envisager l'étude des comportement collectifs, y compris de modèles de la communication.

II.C.2 Modèles réalisés dans l'équipe d'Edelman

Les modèles réalisés dans l'équipe de Gerald Edelman, inspirés de sa théorie intégrée (cf. II.A.2), représentent un compromis intéressant entre l'autonomie (II.C.1) et les capacités cognitives plus développées (II.C.3). L'idée de groupe neuronal est remplacée par des « unités » qui possèdent de nombreuses connections activatrices et inhibitrices, et une fonction de transfert. Ces réseaux ont néanmoins une évolution très différente des réseaux de neurones (Reeke, Sporns et Edelman 1989).

II.C.2.i La série des *Darwin*

II.C.2.i.a Darwin I

Le premier modèle proposé par Edelman (1981) est un « automate darwinien de reconnaissance » qui évoque plus les classifieurs génétiques (Holland et Reitman 1978) que les modèles du système nerveux. Edelman avoue avoir sensiblement imité le système immunitaire[19]. L'automate est structuré en deux couches : après l'entrée *S* (source), viennent des modules de reconnaissance *R* ou « reconnaisseurs », puis des *R-de-R* ou « reconnaisseurs de reconnaisseurs ». On rencontre donc l'idée de catégoriser l'activité de certaines cartes à l'aide d'autres cartes, bien que l'absence de topologie ne permettent pas de parler ici d'une cartographie de l'entrée par les modules *R*, ou de ces modules *R* par les modules *R-de-R*.

La réalisation consiste en une population de chromosomes représentant des groupes neuronaux qui doivent reconnaître une famille d'autres chromosomes, les

[19] Le sélectionnisme est un dénominateur commun à la fois de l'immunologie moderne et du « darwinisme neuronal ».

signaux. Les populations qui reconnaissent efficacement sont favorisées. Il y a ainsi adaptation de la population initiale aux quatre classes de signaux reçus, et cela sans aucune connaissance préalable. On rencontre une suggestion analogue pour la catégorisation par les *R-de-R*, mais aucune précision sur les connexions réentrantes.

II.C.2.i.b Darwin II

Il y a un saut qualitatif très net entre *Darwin I* et *Darwin II* : ce dernier met en place des principes et des méthodes qui seront réutilisés dans *Darwin III* et *IV*. *Darwin II* (Edelman et Reeke 1982) vise à exemplifier la notion de couple de classification (cf. II.A.2.i.c ci-dessus). *Darwin II* a pour tâche la reconnaissance de lettres tracées sur un quadrillage, grâce à la combinaison réentrante de deux modalités, projetées dans des cartes asymétriques. L'analogie proposée par Edelman (1989:48 sq.) montre la « vue » d'un côté, chargée de reconnaître les caractéristiques de détail de la lettre examinée, et le « suivi de contour par le toucher » de l'autre côté, chargé de reconnaître les caractéristiques générales de la classe. Dans la mesure où nous retrouverons cette inspiration pour notre modèle de la référence dans un agent (cf. VI.D.3), il est utile de la détailler ici.

Pour la première modalité, la lettre projetée sur la matrice (de pixels) d'entrée, est analysée en chaque point par des « reconnaisseurs » qui détectent la présence locale de certains traits, comme le caractère rectiligne, un angle et son sens, etc. Les unités de la deuxième couche, ou *R-de-R*, reconnaissent des configurations de la première couche, à travers un grand nombre de connexions aléatoirement distribuées, excitatrices ou inhibitrices. Ainsi, ces unités répondent aux « configurations globales de propriétés locales », qui ne suffisent cependant pas, selon Edelman, à déterminer la catégorie de la forme présentée. En tout, cas on reconnaît dans cette hiérarchie les idées de *Darwin I* (et cette branche est nommée *Darwin*).

La deuxième branche (nommée *Wallace*), est chargée de déterminer les traits non locaux de la lettre présentée, analogues selon Edelman aux traits détectés par le toucher, quand on suit le contour d'un objet. Cette branche reconnaît par exemple la topologie des intersections. Elle aussi possède deux niveaux, mais le premier est simulé afin de reproduire les caractéristiques attendues du mouvement de suivi de contour ; le second est analogue à celle de l'autre branche.

Les deux répertoires d'unités *R-de-R* sont connectés de façon réentrante : l'activation simultanée de deux unités renforce leur connexion. Cela permet ensuite, après entraînement, de pouvoir associer une classe , à savoir un patron d'activation du répertoire *R-de-R* de *Wallace*, à une forme présentée en entrée. L'idée sous-jacente est que les *R-de-R* de *Darwin* , une fois activés, vont à leur tour activer des *R-de-R* de *Wallace*, qui vont activer d'autres *R-de-R* de *Darwin*, jusqu'à une certaine stabilité. La réponse du système est ainsi constituée par le patron d'activation dans les deux répertoires *R-de-R*, car il n'y a pas d'autre « sortie » du système.

II.C.2.i.c Darwin III

Après avoir étudié la validité de la réentrance et de la catégorisation dans *Darwin II*, une utilisation à plus grande échelle de ces idées est tentée avec *Darwin III*. Les réactions comportementales du système mettent en évidence ses états internes. *Darwin III* regroupe 43 répertoires et plus de 50 000 unités ; la description détaillée du système (Reeke, Finkel, Sporns et Edelman 1990) est longue, mais explicite et convaincante, sauf en ce qui concerne la façon dont les auteurs ont pu régler les nombreux paramètres du système.

Nous décrirons la structure et le comportement des effecteurs de *Darwin III,* simulés par un programme. Le système dispose d'un œil avec une zone centrale et une zone périphérique, qu'il apprend à diriger vers tout objet apparaissant dans le plan scruté, qui sera catégorisé selon la modalité visuelle. Le système possède un bras avec deux articulations planes, qu'il apprend à contrôler afin d'atteindre tout point du plan. Les connexions sensori-motrices ont été réalisées de façon à ce que *Darwin III* désire toucher le contour des objets qu'il voit, en déplaçant son bras. Cela fournit une deuxième modalité dans laquelle catégoriser l'objet, comme le « toucher » de *Darwin II*, et permet d'établir un couple de classification.

Plus discutable est le réflexe de répulsion prédéfini, pour une certaine classification de l'objet (par exemple rayé et bosselé). Ce réflexe manifeste par le comportement certains états internes du système. Les *valeurs* sont donc codées dans les connexions, et cela paraît difficilement extensible pour des valeurs ou buts plus complexes, dépendant de perceptions plus variées. Mais *Darwin III* reste une simulation fort convaincante : elle est fondée sur une théorie descriptive du système nerveux, et ses résultats sont conformes aux prévisions, malgré la complexité de l'architecture.

II.C.2.i.d Darwin IV

Darwin IV utilise une structure de contrôle proche de *Darwin III*, mais le système est muni de capteurs/effecteurs réels, dans un environnement réel (Edelman, Reeke, Gall, Tononi, Williams et Sporns 1992, Reeke, Sporns, Gall, Tononi et Edelman 1994). Il s'agit de roues et moteurs, de capteurs photométriques, et d'une pince articulée ; le robot se déplace dans une pièce, mais reste relié à un ordinateur. Les tâches sont semblables à celles de *Darwin III* : après un apprentissage qui consiste à repérer des blocs colorés, à s'en approcher et à les saisir, un réflexe de « ramassage » incite *Darwin IV* à grouper certains blocs dans un tas.

II.C.2.ii Le modèle *RCI* (« *Reentrant Cortical Integration* »)

Ce modèle de la vision est quelque peu distinct de la série des *Darwin* (Edelman 1989:72-86, Reeke, Finkel, Sporns et Edelman 1990:627-647). Les unités de base, que l'on hésite à assimiler à des GN, possèdent une fonction numérique de transfert, appliquée à des entrées activatrices ou inhibitrices. Ces unités sont groupées en sous-répertoires ou matrices rectangulaires, isomorphes au champ visuel d'entrée ; il y a 1024 unités dans une telle matrice (32×32). Ces matrices sont groupées par 8 dans des

répertoires, chacune correspondant à une orientation donnée du plan (0°, 45°, 90°,…, 315°). Chaque répertoire a une fonction de traitement de l'image, qu'il doit exercer dans chacune des directions correspondant aux sous-répertoires, telle la détection de contours. Il y a 16 répertoires, correspondant à 16 traitements, groupés en trois « aires », portant des noms de régions corticales identifiées chez le singe[20].

Ces matrices sont connectées point à point, ce qui permet l'injection directe d'informations d'un niveau vers un autre ; toutefois, cela montre qu'il n'y a pas vraiment d'abstraction de l'information. Les répertoires connectés forment une arborescence, partant des fonctions les plus simples, à la racine, vers les fonctions ou traitements plus complexes, au sommet. Il y a quelques connexions réentrantes (ici descendantes), et deux répertoires chargés de la résolution des conflits.

Il n'y a pas vraiment de résultat produit : l'expérimentateur observe et interprète l'état des répertoires. Le système analyse et interprète des croquis avec certains contours cachés, ou des interruptions de lignes, etc. Il montre notamment comment la réentrance rend compte des illusions du triangle de Kanisza, et déduit de la présence chez l'homme de ces illusions la réalité des connexions réentrantes.

II.C.2.iii Orientations plus récentes

Les travaux plus récents de l'équipe d'Edelman semblent se concentrer sur des systèmes moins intégrés que la série des *Darwin*. Notamment, deux articles étudient une mesure de la complexité d'une carte corticale, ou de sa structuration (Tononi, Sporns et Edelman 1994, 1996). Aussi, une mesure de l'information mutuelle entre cartes est proposée, et sa variation lorsqu'une carte catégorise des signaux du monde extérieur. Il ne semble donc pas que la direction privilégiée soit celle des systèmes autonomes, malgré l'expérience avec *Darwin IV*, mais des travaux sur les propriétés plus locales des cartes – dont le détail vaut d'être exploré.

II.C.3 Autres systèmes cognitifs

La troisième catégorie de systèmes isolés concerne des systèmes qui se rapprochent du paradigme des agents : ils possèdent souvent des perceptions et des actions, mais leur position par rapport à l'environnement est peu autonome. Les aspects qui nous intéressent sont précisément les capacités « linguistiques » de ces systèmes, qui sont en général absentes des agents autonomes vus jusqu'ici.

On pourra bien sûr objecter que presque tout programme de TAL peut être vu comme un agent : en fait, nous nous intéressons ici aux *agents linguistiques*, i.e. ceux qui tentent de reproduire, de près ou de loin, la capacité linguistique humaine. Ainsi, un analyseur syntaxique ne peut être vu ici comme un agent linguistique. Il y a en revanche d'autres systèmes que nous n'aborderons pas, il s'agit des programmes de dialogue homme-machine encapsulés dans des « agents » pourvus de buts et capables d'action, tel COLLAGEN (Rich et Sidner 1997). Ces programmes utilisent souvent pour le langage

[20] Il y a donc au total 129 sous-répertoires (16×8 + 1, le dernier étant celui d'entrée) et 132 096 unités.

naturel les techniques symboliques classiques, qui interagissent peu avec la structure d'agent.

Nous nous restreindrons à quelque exemples de systèmes utilisant les réseaux de neurones pour différentes tâches liées au langage naturel. Ces exemples adoptent l'approche dite sub- ou neuro-symbolique (cf. I.C.2.ii). Un exemple qui aurait pu figurer ici est le modèle connexionniste intégrant lexique, syntaxe, *scripts* et mémoire proposé par Risto Miikulainen (1993), mais nous n'avons pu nous procurer ce livre important.

II.C.3.i Traitement de données séquentielles (Elman)

Jeffrey Elman (1990) a proposé une architecture connexionniste récurrente, contenant une couche dite « d'état », qui permet de « retenir » l'état du réseau à l'instant t_{n-1} afin de le répercuter sur l'état à l'instant t_n. Cela permet à un réseau d'apprendre des séquences de données présentées dans la couche d'entrée. Elman applique ce modèle à la généralisation lexico-syntaxique par un réseau. L'une des expériences consiste à présenter au réseau de courtes phrases, générées automatiquement, de la forme sujet/verbe, ou sujet/verbe/objet. Les couches d'entrée et de sortie contiennent 31 neurones artificiels, chacun des 23 mots de l'expérience (noms et verbes) étant codé par une unité. Les phrases sont données les unes après les autres, et la tâche du réseau est d'apprendre l'ordre des mots, afin de produire en sortie le mot prévu à t_{n+1} ; l'apprentissage consiste simplement en une suite suffisamment longue, sans utiliser d'algorithme qui compare la sortie avec le résultat désiré et modifie les poids des connexions (rétropropagation). La séquence d'apprentissage d'environ 27 000 mots (10 000 phrases) est présentée six fois avant de mesurer les résultats.

On constate alors une erreur de prévision assez faible[21]. Qui plus est, la couche interne fournit une sorte de « représentation » des mots entrés, structurée en fonction des cooccurrences de ces mots dans le corpus. Des analyses par clusterisation de l'activation des unités internes montrent des similarités accrues pour les noms d'une part et pour les verbes de l'autre. En fonction des verbes qui peuvent les suivre, les noms sont aussi regroupés par classes sémantiques (animaux, humains, nourriture, cassables, etc.), et réciproquement, les verbes se classent aussi, mais selon le caractère transitif, intransitif ou mixte ; ces classements sont approximatifs, car il reposent sur une ressemblance approximative des patrons d'activation.

Plusieurs autres exemples sont examinés. Ainsi, le remplacement d'un mot du corpus d'entrée par un mot nouveau aboutit à un classement de ce dernier, dans la représentation de la couche cachée, dans un cluster assez proche du mot remplacé. Enfin, une analyse par clusters de l'activation de la couche cachée pour chacune des phrases du corpus montre que ces phrases sont distinguables, mais très proches ; si l'on regroupe les

[21] L'erreur de prévision est en moyenne de 0,88 unités de différence (entre l'activation réelle et celle attendue, pour la couche de sortie). En réalité, la prévision du mot suivant n'est pas possible exactement, seulement sa catégorie. C'est pourquoi l'on doit comparer la sortie, en moyenne, avec la distribution de probabilités conditionnelles déterminée à partir du corpus ; l'erreur est alors de 0,053.

phrases qui contiennent un même mot, on obtient une représentation analogue à celles des classes sémantiques précédentes, mais naturellement bien plus détaillée. Selon Elman, ces deux aspects fournissent la base de la distinction type/instance, qui serait ainsi ancrée dans le réseau. Cette interprétation paraît toutefois contestable, étant donné la simplicité du modèle : il s'agit plutôt d'une lecture très spéciale des résultats.

II.C.3.ii Association images/phrases (Nenov et Dyer)

Le système DETE (1988b, Nenov et Dyer 1988a) apprend à associer des scènes visuelles se déroulant sur une matrice d'entrée avec des séquences lexicales ou phrases, chaque mot étant codé par un neurone artificiel[22]. L'implémentation utilise, de façon très originale, des neurones artificiels qui s'échangent des potentiels d'action, dans un dynamique temporelle et non les « niveaux d'activation » classiques. Il semble toutefois que toutes les possibilités de la dynamique ne soient pas utilisées, par exemple les régimes oscillatoires, la synchronisation de phase, etc.

DETE est capable de catégoriser des scènes simples, où un objet circulaire ou carré se déplace dans une direction, heurte ou non les murs et rebondit. Parallèlement, des phrases approximativement synchronisées avec les scènes lui sont présentées ('balle bouge haut heurte bouge bas'). Le système semble capable d'apprendre ces phrases pour pouvoir ensuite décrire lui-même, à l'aide de phrases, une scène visuelle qui lui est présentée. D'autres structures internes sont décrites, qui devraient permettre la généralisation et aussi l'« imagination » à partir de mots.

II.C.3.iii Apprentissage d'un système de prépositions (Regier)

D'inspiration analogue au modèle DETE, mais avec des ambitions plus mesurées, le système conçu par Terry Regier (1996) vise à apprendre à catégoriser des scènes spatiales, sur la base d'instances représentant un nombre varié de prépositions. Par exemple, il peut apprendre le système de prépositions spatiales de l'anglais (*above*, *below*, *left_of*, *right*_of, *around*, *in*, *on*, *out_of*, *through*, *over*), qui concernent aussi bien la position que le déplacement. Après des exemples de chaque cas, le programme généralise et catégorise correctement des cas nouveaux. Plusieurs systèmes de prépositions sont apprenables (cinq langues testées) ; pour chaque système, le programme met en évidence les cas de recouvrement entre prépositions.

L'architecture du système est tout entière orientée vers le traitement de séquences d'images ; plusieurs « cartes », analogues par exemple à celles du modèle *RCI* d'Edelman (cf. II.C.2.ii), détectent différents traits dans la matrice d'entrée. La séquentialité est traitée à l'aide d'un réseau récurrent comme celui d'Elman. Deux objets figurent dans chaque séquences d'images projetée, un repère (*landmark*) et un mobile (*trajector*), ce qui est inspiré des descriptions de la grammaire cognitive. Enfin, dans la couche de sortie, chaque neurone correspond à l'une des prépositions devant être

[22] Le système a fait l'objet de la thèse de Nenov (Nenov 1991). Notre description est fondée sur (Dyer 1991:76-80), seul document que nous avons pu consulter.

apprises. L'architecture d'ensemble est connexionniste, et préalablement structurée ; elle s'inscrit dans le paradigme du *connexionnisme structuré*, qui consiste à combiner manuellement des sous-réseaux de neurones classiques afin d'accélérer l'évolution d'un grand réseau non structuré vers un réseau adapté au problème traité.

En conclusion de ces trois réalisations (Elman, Nenov & Dyer, Regier) on comprend mieux leur dimension complémentaire par rapport aux animats du II.C.1 ci-dessus. Ces systèmes présentent des traits adaptatifs, et neuro-mimétiques, mais qui sont appliqués exclusivement au traitement de données (pseudo)linguistiques, ou à la réception passive des états de l'environnement. Les trois systèmes, malgré leurs potentialités, peuvent difficilement être considérés comme des agents, puisqu'ils n'ont ni effecteurs ni besoins, et n'évoluent pas dans un environnement. En revanche, parmi les systèmes traitant la langue naturelle en général, ces trois systèmes semblent plus proches d'une intégration à un agent bénéficiant déjà d'une structure neuronale pour la perception et l'action. Il resterait bien sûr à redéfinir la tâche de ces systèmes, afin de la relier à la « survie » d'un agent dans son environnement.

II.D Interactions dans les sociétés d'agents

Nous arrivons maintenant aux réalisations touchant directement le domaine que nous défendons, à savoir l'utilisation de systèmes multi-agent pour simuler certains phénomènes liés au langage. Avant d'examiner ces réalisations (II.D.2), il est nécessaire de les distinguer d'un domaine très voisin, mais obéissant à des motivations différentes : les protocoles de communication formels dans les systèmes multi-agents « classiques » (II.D.1). Ces protocoles abstraits et symboliques sont conçus pour être précis et efficaces (mais non compréhensibles par des humains), et ils sont utilisés dans les systèmes multi-agents visant la résolution collective de problèmes.

II.D.1 Logique et langage dans les systèmes multi-agents

II.D.1.i Formalisations de la communication dans les SMA

La communication des agents est étudiée ici à un niveau conceptuel ; l'implémentation physique des processus est indifférente (protocoles de bas niveau comme TCP/IP, ou bien HTTP, CORBA, etc.). Ce qui compte, c'est la nature de l'information transmise entre agents logiciels, et son influence sur les actions.

II.D.1.i.a Les actes de langage

La théorie des actes de langage (Austin 1962, Searle 1969) explique le fonctionnement des messages dans la communication humaine, ou encore leur rôle *illocutoire*, à l'aide de deux éléments : la force illocutoire ou le prédicat performatif (*affirmer*, *ordonner*, *promettre*, *informer*, *interroger*, etc.) et le contenu propositionnel (l'assertion affirmée, l'ordre donné, la promesse, etc.). Ces considérations (simplifiées ici à l'extrême) ont attiré l'intérêt de la communauté multi-agent car elles permettent

d'aborder de façon formelle, donc programmable, la communication efficace entre agents, orientée vers l'action.

Pour ce faire, une sémantique des actes de langage est recherchée (Vandervecken 1990a, 1990b), notamment la sémantique des prédicats performatifs. Un essai relativement connu est le langage KQML[23], en réalité une liste d'environs quarante performatifs élémentaires, accompagnés d'une description verbale de leur sens. L'ambiguïté, l'incomplétude et la faiblesse de la formalisation de KQML ont été critiquées notamment par (Cohen et Levesque 1995).

II.D.1.i.b Les modèles de conversations

La conception de systèmes multi-agent a fait naître la nécessité d'une formalisation des conversations (échanges de messages) possibles entre agents, afin de régler de manière absolument exacte leur interaction.

- *Formalismes indépendants*

La modélisation abstraite des conversations, dans leur enchaînement formel, peut être décrite par des automates à états finis, comme dans le modèle de (Winograd et Flores 1986), ou bien par des réseaux de Pétri (Ferber 1995:348-358). Ce dernier travail propose également une sémantique de quelques actes de langage.

- *Formalismes fondés sur une sémantique des actes de langage*

Outre leur critique de KQML, Cohen et Levesque (1995) ont proposé un modèle d'agent rationnel fondé sur le formalisme des mondes possibles, et débouchant sur une sémantique des actes illocutoires. Une tentative ayant le même but, mais reprenant les standards KQML, a été proposée par Labrou et Finin (1997). Ce travail fournit une sémantique opératoire des prédicats performatifs, en termes des pré et post-conditions, et de changement de l'état cognitif des agents. Bien évidemment, aucune considération sur le *contenu* propositionnel ne rentre en jeu ; celui-ci est représenté par une variable, comme dans *believe*(*agentX*, *propositionY*), et de même pour *know*(…,…), *want*(…,…) , *intend*(…,…). Les auteurs spécifient ensuite des protocoles de conversation réalisant ces illocutions, à l'aide de grammaires formelles. Signalons aussi un autre langage pour la communication entre agents, l'ACL proposé par la FIPA[24].

II.D.1.ii Comparaison avec les modélisations du langage humain

II.D.1.ii.a L'exemple du projet PVS (LIMSI)

PVS est une architecture multi-agent générique (appliquée à la gestion de projets) développée dans notre groupe. La communication entre les agents (Guichard et Briffault 1998) utilise leurs capacités de raisonnement symbolique, à l'aide de techniques d'IA classique pour modéliser les croyances, les désirs, les intentions. Les messages se

[23] « Knowledge Query and Manipulation Language » (langage de requête et manipulation de connaissances), projet de la DARPA aux États-Unis.

[24] ACL : Agent Communication Language ; FIPA : The Foundation for Intelligent Physical Agents.

composent d'un prédicat performatif et d'un contenu, exprimé à l'aide de graphes conceptuels (Sowa 1984, 1992).

Les auteurs pensent reproduire les « niveaux » traditionnels de la langue : à la syntaxe correspondraient les différents champs du message (*destinataire*, *envoyeur*, *performatif*, *contenu*, etc.), au lexique les noms des performatifs (sans le contenu). La sémantique des actes, dans l'esprit de celle proposée par (Labrou et Finin 1997), définit la modification de l'état cognitif de l'agent après un acte de langage ; cela ne dépend pas du contenu propositionnel, seulement des croyances, buts, etc.[25] La pragmatique semblerait être reliée au contexte d'énonciation (*sic*). On comprend donc sur cet exemple que l'analogie entre la communication formelle entre ces agents et la simulation d'une langue est très superficielle.

II.D.1.ii.b Différence de buts

Les buts dans les deux cas sont presque opposés. Dans les systèmes multi-agents classiques, la communication entre agents doit être la plus transparente et exacte possible, et sert globalement à la résolution de problèmes par le système (à savoir sa tâche). L'échange de messages doit être prévisible par le concepteur, et le code utilisé ne doit pas évoluer. La codification des messages est conçue indépendamment de la langue, avec l'efficacité comme unique souci.

Lorsque l'on tente de modéliser des phénomènes linguistiques dans un groupe d'agents, le but est d'illustrer certaines propriétés du langage (collectives ou individuelles). Le code utilisé est en général évolutif, parfois ambigu, et sujet à conventions entre les agents. On étudie alors sa dynamique, éventuellement son lien avec le système représentationnel des agents, ou sa capacité à augmenter la survie de ceux-ci. Ce dernier point n'est souvent pas un but de l'expérience, mais un moyen de faire émerger des phénomènes complexes.

II.D.1.ii.c Rapport éloigné avec le langage

Les rapports existent, certes, entre ces deux approches de la *communication*, d'où un possible intérêt des protocoles formels pour la modélisation du langage.

- *Une origine commune ?*

La recherche d'un code de communication efficace s'est naturellement tournée vers le langage ; la théorie des actes de langage a d'abord été conçue et employée en philosophie du langage, son application au multi-agent étant plus tardive. Toutefois, il faut remarquer l'importance des théories du choix rationnel, et de croyances/désirs/intentions, qui éloignent du langage les SMA formels.

- *Un point de rencontre : la pragmatique*

On peut dire que les bases communes des deux approches se retrouvent dans la pragmatique, plus particulièrement l'étude de l'interprétation/génération linguistique en

[25] Évidemment, aucun paradoxe n'est possible, comme « croire qu'on ne croit pas P », puisque le domaine des propositions crues ne contient pas d'attitude intentionnelle.

contexte ; du contexte font partie aussi les univers cognitifs des locuteurs. Les rapports entre pragmatique et linguistique ne sont pas simples, mais on note en général qu'aucun des deux domaines n'est inclus dans l'autre, et que leur intersection est assez réduite (Moeschler et Reboul 1994:17-41).

- *Rôle réduit de la linguistique dans la formalisation*

À partir du terrain d'entente pragmatique, les deux approches présentent des tendances divergentes. On peut dire que l'étude des protocoles formels évite tout appel à la linguistique (lexique, syntaxe, sémantique du lexique) pour n'introduire aucune « imperfection » et construire un code « parfait ». Au contraire, la simulation de la langue recherche les phénomènes de fluctuation, ambiguïté, convention, convaincue qu'elle est que celles-ci fournissent au langage ses capacités adaptatives essentielles et sa robustesse, qui semblent faire défaut aux programmes de TAL.

II.D.2 Modélisation multi-agent des échanges linguistiques

Le domaine qui nous intéresse a vu une multiplication des travaux ces dernières années, signe que l'évolution du langage, et en particulier les modèles de son émergence, peuvent être sujet d'étude et de débat *scientifique*. Ce développement fait suite aussi aux travaux sur les sociétés d'agents et en particulier sur la vie artificielle, dont l'essor date de la fin des années '80. Témoin de l'intérêt pour les modèles de l'émergence du langage, on notera la récente série des conférences *Evolution of Language* (1996, 1998, 2000)[26], ainsi que la revue *Evolution of Communication*.

Le premier numéro de cette revue contient une remarquable synthèse sur les modèles multi-agent de l'évolution du langage (Steels 1997c) – c'est pourquoi nous nous limiterons ci-dessous à un bref aperçu du domaine et à des évaluations critiques. Dans l'article cité, Steels expose le principe des « systèmes adaptatifs complexes » qui sous-tend ces simulations, et classe celles-ci selon leur principe d'évolution : purement génétique, par adaptation (auto-organisation), ou par assimilation génétique (couplage des deux premières). Steels propose aussi une description de ces systèmes selon six critères, qui pourraient devenir un standard : (1) nature de la population, (2) structure innée des agents, (3) interaction linguistique, (4) adaptation des agents, (5) source de nouveauté, (6) interaction génétique.

Par rapport aux exemples précédents (II.C), les agents utilisés dans les présents modèles ne visent pas directement l'autonomie ou l'adaptabilité des uns (II.C.1), ni les facultés cognitives ou linguistiques des autres (II.C.3). Ces agents empruntent quelque chose aux *deux* visées, et leur ajoutent une faculté de communication. Les trois fonctions – autonomie, perception/action, communication – auront des proportions variées, et *aucune ne sera négligeable*. Nous allons étudier quelques exemples classés par ordre de complexité, finissant par les exemplaires travaux de Luc Steels.

[26] Le recueil (Hurford, Studdert-Kennedy et Knight 1998) a été tiré des actes de la première édition. La troisième édition aura lieu à Paris en avril 2000 (http://www.infres.enst.fr/confs/evolang/).

II.D.2.i Émergence et évolution de systèmes de signalisation

Des capacités très élémentaires de communication peuvent donner naissance, dans un système multi-agent, à des comportement complexes et « intelligents ». On peut ainsi citer les modèles de sociétés de fourmis, qui utilisent des agents réactifs communiquant à l'aide de marques posées dans l'environnement ou « phéromones » (Deneubourg, Goss, Sendova-Franks, Detrain et Chrétien 1991, Drogoul et Ferber 1994). Toutefois, cette communication semble trop éloignée du langage.

L'« éthologie synthétique » proposée par MacLennan (1992), ainsi que les expériences menées avec J. Crumpton (Crumpton 1994, MacLennan 1990), constituent une des premières simulations mettant en évidence l'émergence d'un « langage » qui permet à des individus de mieux survivre. Toutefois, les éléments de l'expérience sont hautement schématiques : agents fixés dans un environnement circulaire, mécanisme rigide d'observation d'une « scène » commune et de production d'un signal, système de récompense très élémentaire.

L'expérience de Werner et Dyer (1992) apporte plusieurs éléments réalistes par rapport à la précédente. Ainsi, les agents peuvent se déplacer dans un environnement, et le code de signaux permet aux « femelles » de guider les « mâles » vers elles ; le bénéfice du code n'est donc plus individuel, mais se répercute sur les générations suivantes. Toutefois, la structure des agents reste très élémentaire et relativement figée. Ceux-ci sont constitués uniquement d'un micro-réseau de neurones (3 couches à 3-4 unités) responsables des associations « mot »↔déplacement (pour les mâles) et « perception d'un mâle »↔« mot » (pour les femelles).

Une expérience plus complexe, mais de facture analogue, a été conçue par Cangelosi et Parisi (1996). La fonction des signaux est d'améliorer, grâce à l'expérience collective, l'identification par les agents de « champignons vénéneux ». La structure neuronale des agents (3 couches : 14, 5, 5) est plus complexe, et l'analyse de la dynamique du code de signalisation plus détaillée que dans l'expérience précédente.

II.D.2.ii Une expérience de couplage entre « sémantique » et « grammaire »

John Batali (1998) propose un modèle d'agent dont la structure interne utilise des réseaux récurrents, comme ceux d'Elman (II.C.3.i). Cela lui permet de simuler la production de messages séquentiels par les agents (chaînes de caractères) et leur interprétation séquentielle. Les agents possèdent un « vecteur sémantique », c'est-à-dire une couche de 10 neurones représentant les états conceptuels, interprétés sous la forme d'un prédicat – 2^6 possibilités – et d'un pronom déictique de personne – 2^4 possibilités, chaque bit recevant une interprétation sémantique. Le but des agents est de transmettre aux autres, lors des dialogues, leur état conceptuel ; pour ce faire, des séances d'apprentissage sont organisées, dans lesquelles les récepteurs ajustent les poids de leur réseau en fonction du sens donné qu'ils *doivent* trouver.

Un point intéressant est l'émergence du couplage entre la « sémantique », i.e. le sens qu'un agent doit transmettre, et la « grammaire », i.e. la séquence de caractères, de longueur arbitraire, que celui-ci décide de transmettre. Ce couplage est réalisé grâce au

réseau récurrent, qui permet à l'émetteur d'interpréter son propre message en même temps qu'il le produit ; ainsi, il pourra apprendre à produire la séquence de caractères qu'il attendrait, lui, pour l'état conceptuel respectif. Évidemment, nous avons affaire ici plus à une *interprétation* du modèle en termes anthropomorphiques – de même que l'on attribue à la séquence de caractères des propriétés « grammaticales ». Dans un article à paraître, Batali (1999) aborde la dimension récursive des grammaires, mais à l'aide d'un modèle symbolique, fondé sur le calcul des prédicats du premier ordre.

II.D.2.iii Dynamique des grammaires formelles

Sur un plan proche de la conception standard des grammaires formelles, Hashimoto et Ikegami (1998b, 1996) étudient l'évolution des grammaires dans une population d'agents en interaction. Ces agents ont la capacité de former et respectivement analyser des chaînes de caractères (appelées « mots ») en utilisant des règles grammaticales formelles. Les auteurs mettent en évidence une évolution vers un enrichissement de la grammaire (non seulement des règles, mais de son *type*), ce qui permet aux agents de produire et comprendre des mots de plus en plus variés.

Le principal problème de ces simulations, pour le domaine qui nous intéresse, est leur caractère purement formel : il n'est question ni de situation dans un environnement, ni de perception/action, ni d'une quelconque forme de contenu. Selon Takashi Hashimoto (communication personnelle), cela correspondrait à une conception de la langue comme système autonome de symboles (au sens structuraliste) qui prennent en grande partie leur sens les uns par opposition aux autres – ceci étant vrai bien sûr du système lexical (« tokens »), mais aussi du système des concepts si ceux-ci reposent sur des conventions systémiques.

Ce point de vue est illustré par les travaux de Hashimoto (1998a) sur l'émergence de relations entre mots à partir de l'analyse statistique de leur contexte d'utilisation par les agents ; ces relations sont comparées à des relations sémantiques. Toutefois, une telle conception de la langue comme système de symboles ne doit pas passer sous silence *l'ancrage* des signifiés au niveau individuel, comme le montrent Victorri et Cazoulat (1993).

II.D.2.iv Les travaux de Luc Steels

La somme de travaux réalisés par Luc Steels et son équipe[27] aborde l'ensemble des tendances esquissées jusqu'ici, du lexique à la syntaxe, selon une remarquable progression. On notera d'abord la préoccupation constante de valider expérimentalement les modèles proposés, par la mise en évidence de comportements concrets de robots ou d'agents simulés (II.D.2.iv.a), et l'étude quantitative et/ou graphique des propriétés linguistiques émergentes. Après un aperçu de ces travaux (II.D.2.iv.b), nous tenterons de dégager leur principaux apports (II.D.2.iv.c).

[27] Ou plutôt ses équipes, celle du Laboratoire d'IA de la Vrije Universiteit Brussel (VUB), et celle, plus récente, du Sony Computer Science Laboratory (CSL), Paris.

II.D.2.iv.a L'environnement expérimental multi-agent : robotique, simulation

L'environnement robotique utilisé par Steels à Bruxelles a été développé par les travaux sur le contrôle sensori-moteur et l'interaction sociale des agents physiques, lorsqu'ils sont en compétition pour des « ressources », par exemple la lumière (Steels 1994a, 1994b). Ces robots maintenant bien maîtrisés sont munis de nombreux capteurs de pression et de luminosité ; leur étude montre bien que le passage de l'étude des simulations à leur réalisation physique est loin d'être trivial (capteurs et moteurs imprécis, ressources embarquées limitées).

Afin de disposer d'une véritable perception visuelle grâce à une caméra, des robots non mobiles ont été conçus, pouvant uniquement orienter cette caméra. Groupés par deux, dans deux expériences intitulées « Talking Heads »[28], ces robots sont situés soit dans l'environnement robotique précédent (Steels 1997b), soit devant un tableau magnétique muni d'objets (Steels et Kaplan 1999).

Les simulations abstraites (non robotiques) programmées ont quant à elles deux fonctions. La première semble être le test des modèles avant leur implémentation sur robots physiques – tel le programme LISP étudiant l'émergence du lexique (Steels 1996c), ou les explorations de la syntaxe dans (Steels 1998a). Mais ces simulation servent surtout à étudier des modèles à grande échelle (de temps ou nombre d'agents) de l'évolution linguistique ; c'est le cas de la plate-forme *Babel* développée par Angus McIntyre (1998) et utilisée dans de nombreuses études.

II.D.2.iv.b Études expérimentales : catégorisation, phonologie, lexique, syntaxe

Les expériences de Luc Steels abordent l'émergence et l'évolution de plusieurs aspects de la communication langagière, dans une approche *incrémentale*. On peut ainsi constater que la syntaxe est devenue à présent un problème central, après le lexique et la sémantique, comme le montre le bref résumé suivant.

- La catégorisation perceptive des robots consiste en une division de l'intervalle de valeurs possibles de chaque paramètre perceptif en segments, qui peuvent à leur tour être divisés (dichotomie). Ce mécanisme convient à la fois aux capteurs rudimentaires des robots mobiles (intensité lumineuse visible ou infrarouge, etc.) et aux caméras (position horizontale/verticale, intensité R/G/B). Le développement de telles catégories se fait par autant de dichotomies d'intervalles que nécessaire (Steels 1996a). Ce mécanisme de catégorisation semble être conservé dans toutes les simulations ultérieures, parmi lesquelles on peut citer l'étude de l'évolution des catégories elles-mêmes (Steels 1997a).

- Des études acoustiques de l'émergence de conventions phonologiques ont été réalisées (Boer 1997) ; dans la mesure où celles-ci ne concernent que le système

[28] Têtes parlantes : http://talking-heads.csl.sony.fr .

vocalique, elles n'ont pas encore été intégrées aux expériences suivantes, qui utilisent des « mots » symboliques de la forme *consonne+voyelle*.

- La simulation symbolique de l'émergence d'un lexique est une des expériences les plus frappantes : les robots donnent des noms aux traits catégorisés, et tentent ainsi, à travers des « jeux de nommage », d'identifier des objets du contexte. Le succès à ce jeu permet à certaines conventions lexicales de se répandre (Steels 1996b, Steels 1996c). Les conventions font émerger des catégories distinctives communes, qui peuvent évoluer si l'environnement change (Kaplan 1998).

- Tout aussi frappante est l'intégration de ces simulations dans la société de robots physiques (d'abord sans caméra), qui demande un contrôle mécanique fin des actes de pointage vers les objets, dans les « jeux de langage ». Ces expériences mettent en évidence un *ancrage par association* du lexique dans les perceptions ; elles sont décrites dans (Steels et Vogt 1997) et synthétisées dans (Steels 1998b).
  Des expériences plus récentes utilisent les caméras pour caractériser de façon plus précise les objets de l'environnement. De nombreux agents logiciels peuvent utiliser les caméras tour à tour, mettant ainsi en évidence l'évolution des conventions de nommage dans une population. Deux variantes des « Talking Heads » sont décrites, dans (Steels 1997b) et (Steels et Kaplan 1999), le degré de finition technique de la seconde étant spécialement réussi.

- Les expériences sur la syntaxe sont par comparaison moins avancées : le problème est dans son ensemble plus difficile. Son émergence est modélisée par la formation d'associations ou cadres regroupant plusieurs mots émis ensemble dans une même situation ; ces cadres se généralisent peu à peu, et engendrent en retour des catégories « syntaxiques » pour les mots (Steels 1997b). Des études théoriques plus récentes leur appliquent un formalisme syntaxico-sémantique assez particulier (Steels 1998a).

- Enfin, une tendance plus récente aussi est l'étude de la diachronie (évolution linguistique) notamment par des simulations sur le lexique (Steels et Kaplan 1998).

II.D.2.iv.c Discussion

Cette revue montre bien les apports des travaux de Luc Steels par rapport aux études plus classiques du langage en intelligence artificielle (cf. **Chapitre I**). L'ancrage du langage dans les facultés sensori-motrices, et surtout dans un système à plusieurs locuteurs, est ici pleinement visible. Une approche « bottom-up » est adoptée, dans laquelle toute avancée se construit grâce aux éléments existants, et peut ainsi être intégrée au monde robotique. De l'aveu de Luc Steels (1997c:4, nous traduisons) :

> *Le paradigme des systèmes adaptatifs complexes et la modélisation par agents sont clairement pertinentes pour le langage, en particulier pour deux*

domaines : la compréhension des origines du langage, et celle de la dynamique évolutive des systèmes linguistiques.

Ces deux domaines correspondent effectivement aux préoccupations les plus fédératrices, mais on peut se demander si l'étude de la *diachronie* ne restreint pas le champ de la modélisation multi-agent, en occultant en particulier la question essentielle du *fonctionnement synchronique* du langage. Certes, une « vraie » explication des origines du langage devra contenir une explication de son fonctionnement, mais la réciproque n'est peut-être pas nécessaire. En somme, on peut donner des modèles de *l'évolution* sans éclaircir par là le *fonctionnement*, ce qui ne correspond pas à nos propres visées.

II.D.2.v Notre conception de la modélisation multi-agent du langage

La technique de la modélisation multi-agent est, nous l'avons dit au Chapitre I, un des éléments clé de la « pragmatique simulée ». Les travaux précédents montrent les possibilités actuelles, mais par rapport à eux, *nous mettons l'accent sur quelques points différents.*

- La modélisation multi-agent devrait permettre, à nos yeux, de progresser vers un système informatique doué de capacités linguistiques élaborées, plus réalistes que les capacités actuelles des programmes de TAL. Aborder le problème de l'origine et de l'évolution du langage est seulement une possibilité des modèles multi-agent, et peut aussi représenter une voie pour atteindre notre but plus général[29]. Toutefois, il n'est pas certain que la re-création intégrale d'un phénomène ayant eu une durée notable, fondé sur un support biologique sophistiqué, soit la solution la plus facile pour conférer une capacité linguistique à des machines.

- Toujours dans l'optique d'une application au TAL, on peut remarquer que les conventions de communication des agents ne se rapportent à aucune langue réelle. Cela participe certes de l'autonomie des agents (ou de son illusion), et se justifie lorsque plusieurs items lexicaux, par exemple, dénotent un seul sens : on ne pourrait arbitrairement en attacher un seul à un mot réel. Toutefois, on ne voit pour l'instant pas de tentative pour imposer des contraintes sur le code que l'on fait évoluer, pour adopter les conventions d'une langue existante (lexique, ordre des mots). Nous proposons au Chapitre III une méthode en ce sens, inspirée de la créolisation.

- Que l'on cherche à mettre en évidence l'origine du langage, ou seulement son état actuel, se pose la question de la complexité requise des agents.

  Il nous semble que les travaux de Steels, dans la perspective « bottom-up », considèrent qu'on ne doit pas compliquer excessivement la structure des agents, en tout cas pas avant d'avoir exploité toutes les propriétés d'une structure simple, dans

[29] Comme nous l'avons commenté dans l'évocation des trois cas d'émergence du langage, au II.B.1.

un contexte collectif et dynamique. Les résultats convaincants obtenus tendent à accréditer cette hypothèse, qui se rattache à l'idée que le langage est dû à un « mécanisme cognitif général », plutôt qu'à un « organe du langage ». Toutefois, la plupart des modèles rencontrés sont tous hautement spécifiques à la tâche requise (la communication), et contiennent de nombreuses instructions préprogrammées essentielles aux dialogues et à la mise à jour des connaissances.

Sur ce point, il est probable qu'une imitation plus poussée des propriétés naturelles s'impose, soit comme le fruit de l'évolution d'une population d'animats (cf. les travaux neurogénétiques, II.C.1.ii), soit en la donnant dès la première génération.

- Une structuration plus élaborée des agents devrait leur permettre un meilleur *ancrage individuel* du langage. On remarque en effet que, dans les modèles précédents, l'intégration entre la faculté de langage et les facultés autonomiques de l'agent (perception/valeurs/actions) est assez faible – quand ces dernières existent. La faculté de langage devrait être ancrée aussi bien dans les perceptions (base des catégories, comme dans de nombreux modèles) que dans les valeurs (catégorisation fondée sur la valeur, et en fin de compte *fonction* du langage) et dans l'action (ordonnancement, performatifs). À ce titre, les expériences de Steels avec les robots mobiles et autonomes (Steels et Vogt 1997) semblent plus réalistes que les « Talking Heads », du point de vue de l'intégration perception–action–valeurs–langage.

* * *

On comprend maintenant, au terme de cette revue, l'enchaînement nécessaire des connaissances exploitées, et l'éclairage qu'apportent les études du langage réel, notamment son support humain et son émergence dans l'individu et la collectivité, sur les tentatives de modélisation réalisées jusqu'à présent.

Nos commentaires à propos des modèles informatiques présentés tracent la voie de notre progression future. Nous avons pu préciser quels éléments paraissaient les plus prometteurs pour la modélisation des phénomènes individuels et collectifs ; notre insistance particulière sur *l'ancrage* du langage dans l'individu ne doit pas faire oublier la difficulté du problème, qui explique les solutions relativement simples adoptées dans la modélisation par agents.

Afin de progresser dans cette direction, nous exposerons d'abord notre propre cadre de modélisation des échanges entre agents, que nous appelons « pragmatique simulée ». Deux expériences, à deux niveaux différents de ce cadre, sont ensuite décrites. Elles ne doivent pas faire oublier un problème central des rapports langage/ individu/environnement, celui de la référence, qui occupera notre deuxième partie.