

Chapitre 7

Conclusions et perspectives

*Quand au fond, je n'ai pas voulu faire semblant d'inventer
ce que la tradition m'offrait, quand je ne faisais que le reprendre.
Non que je n'aie rien mis de mien, dans cet ouvrage, au contraire!
Mais on ne possède jamais que ce qu'on a reçu et transforme, que ce
qu'on est devenu, grâce à d'autres ou contre eux.*

André Comte-Sponville, *Petit traité des grandes vertus* [Comte-Sponville, 1995] (p.11)

7.1 Résumé et synthèse

Le but principal de cette thèse était de poser les premières briques d'un scénario général de l'émergence et de l'évolution préhistorique des langues. L'idée n'était pas de remettre en question les principales théories à ce sujet, mais plutôt de les préciser et de souligner certains points importants susceptibles d'avoir joué un rôle significatif dans la construction des langues telles que nous les observons aujourd'hui. Pour ce faire, plusieurs principes ont guidé nos études :

- les aspects systémiques du langage, comme mis en évidence par Saussure il y a plus d'un siècle, ont joué un rôle important et peu étudié dans l'émergence, la complexification et plus généralement l'évolution des langues au cours de la préhistoire ;
- les modèles et les simulations informatiques sont des outils de choix pour l'étude de phénomènes linguistiques du passé, en particulier à cause de l'absence de "fossiles linguistiques" ;
- une approche pluridisciplinaire, reposant non seulement sur la (socio-)linguistique et l'informatique, mais aussi sur l'archéologie, les théories de l'évolution, la paléo-anthropologie (cognitive). . . est le plus sûr moyen de découvrir et d'asseoir les mécanismes de l'émergence et de l'évolution ancienne des langues. Cette approche est même souvent une nécessité pour pallier à la pauvreté des données linguistiques à ce sujet.

Au cours de notre travail, guidés par les principes précédents, nous nous sommes attachés à l'étude des conditions d'émergence des langues à la fois les plus anciennes et les plus contemporaines, ainsi qu'à l'évolution de celles-ci en fonction des structures sociales et démographiques de nos ancêtres. Nous nous sommes centrés en particulier :

- sur la création d'une plate-forme de modélisation puissante et modulaire, au développement incrémental ;
- sur un ensemble de notions systémiques liées au langage : jeu de contraintes, mécanismes de changement, frontières du système. . . ;
- sur la question de la monogenèse ou de la polygenèse des langues, en lien avec l'évolution de notre espèce au sens large : évolution des capacités cognitives ou physiologiques, évolution de la taille de la population et distribution de celle-ci sur le globe, structure sociétale ;
- sur les traversées maritimes vers l'Australie comme moyen de mettre au jour des indices de l'origine des langues modernes ;
- sur la question générale des changements linguistiques : quelle est leur dynamique ? Quels sont les impacts des contraintes sociolinguistiques et internes sur les changements d'un système linguistique ?

Parallèlement à une réflexion plus théorique, des modélisations ont été mises en oeuvre, afin de pouvoir tester certaines hypothèses dans le cadre des contextes et des problématiques précédents. Les principaux résultats ou propositions que nous avons dégagés de ces recherches peuvent être résumés brièvement par les quelques points suivants :

- les langues sont indissociables de l'activité linguistique. Parallèlement à la transmission d'information, l'utilisation sociale du langage et la variabilité qui en découle est une constante au cours de l'histoire des hommes ;

- le développement cognitif et les modifications physiologiques de nos ancêtres ont entraîné un développement de l'espace des possibles linguistiques au cours de la préhistoire lors des événements de spéciation le long de notre lignée. Des structures de plus en plus complexes sont venues enrichir progressivement l'espace des états linguistiques possibles, qui a vu sa diversité s'accroître au cours du temps jusqu'à l'émergence de notre espèce ;
- les items et les structures de l'espace des états possibles sont progressivement apparus dans les langues de nos ancêtres. Il existe une décorrélation entre l'apparition de nouvelles stratégies linguistiques **possibles**, et leur apparition **effective** dans les systèmes linguistiques **réels**. Ces émergences se sont produites selon les cas par monogénèse ou polygénèse, suivant les époques, les structures sociétales de la macro-population préhistorique et des événements contingents. Des contraintes structurelles liées à la composition des langues qui pouvaient accueillir les nouvelles stratégies linguistiques pesaient sur les probabilités d'émergence ;
- grâce aux phénomènes précédents, les états de l'espace des possibles se sont progressivement réfléchis dans l'espace des états réels ; cette réflexion a néanmoins été modulée par les évolutions des structures de la macro-population humaine au cours de la préhistoire ;
- l'étude des traversées maritimes vers l'Australie permet de mettre au jour le développement d'une capacité de **liage représentationnel spatio-temporel** chez les membres de notre espèce, qui sous-tend nombre des révolutions culturelles et techniques observées chez cette dernière ; cette capacité cognitive s'est projetée au niveau linguistique, et les nombreuses stratégies linguistiques apparues en conséquence forment partiellement les caractéristiques des langues modernes par rapport aux langues ancestrales ;
- les traversées maritimes vers l'Australie peuvent être replacées dans le contexte général de l'hypothèse *Out of Africa* et des migrations côtières hors d'Afrique. L'étude des îles Andaman laisse à penser que leurs populations de chasseurs-collecteurs trouvent leurs origines génétiques et linguistiques dans la première migration qui a conduit à la conquête de l'Australie et de la Nouvelle-Guinée ;
- l'impact des contraintes sociales et internes qui pèsent sur un système linguistique peut être évalué assez finement grâce à un modèle et des simulations informatiques qui permettent l'étude des différents paramètres en jeu, en particulier au niveau de la structure des interactions sociales. Il apparaît en particulier que les contraintes internes et sociales peuvent exercer leur influence sur le système de façon **opérationnellement disjointe** ;
- un scénario en quatre phases peut être proposé pour rendre compte de l'évolution de la diversité linguistique au cours de la préhistoire jusqu'aux périodes actuelles.

7.2 Perspectives

7.2.1 Perspectives directes

Développement et ancrage linguistique de l'approche systémique

*Grise, chère amie, est toute théorie,
Et vert l'arbre d'or de la vie.
Méphistophélès, Faust de Goethe*

Il est évident que les propositions formulées au cours de cette thèse manquent de substance, et en particulier de substance linguistique : les caractéristiques typologiques sont abordées de façon extrêmement générale, et peu d'exemples concrets sont utilisés pour ancrer les phénomènes décrits.

Un des développements postérieurs à cette thèse est donc naturellement de renforcer le lien entre propositions théoriques sur les évolutions de la dimension structurelle d'un système linguistique et observations typologiques dans les langues du monde. Notre objectif est de faire coïncider certaines des dynamiques d'évolution que nous avons introduites de façon théorique, comme l'évolution par équilibres ponctués, avec des changements réels des langues du monde. Une telle démarche a été entreprise par certains auteurs, qui proposent par exemple des modèles à base d'attracteurs pour rendre compte d'évolutions syntaxiques réelles [Cooper, 1999].

Utilisation de bases de données phonologiques

Afin de pouvoir mettre à profit des situations linguistiques réelles pour ancrer nos propositions théoriques, nous pouvons envisager deux approches assez différentes. L'une consiste à se focaliser sur des cas d'étude très précis et à tester les théories sur ces quelques manifestations de la réalité linguistique. Une seconde possibilité est le recours à des bases de données typologiques, qui ont l'avantage d'offrir une vue d'ensemble des caractéristiques d'un sous-domaine linguistique.

Les bases de données informatiques les plus utiles pour nos objectifs à l'heure actuelle sont probablement les bases de données phonologiques. Nous disposons ainsi des bases de données UPSID et BDProto, qui contiennent respectivement 451 langues et 101 proto-langues reconstruites. Des collaborations avec d'autres membres du laboratoire *Dynamique du Langage* nous ont déjà permis de mieux apprécier les possibilités d'extraire de l'information de ces corpus de données [Marsico et al., 2002]. Nous disposons également d'une troisième base de systèmes phonologiques, créée par Merritt Ruhlen, et riche d'environ 3000 langues, ce qui constitue un échantillon très important des quelques 6000 langues du monde.

La données des systèmes phonologiques des langues permet de dresser un tableau des fréquences de différents systèmes phonologiques, de différents segments ou encore de différents traits. Il est également possible de travailler sur les co-occurrences de deux ou plusieurs éléments pour tenter de détecter d'éventuelles tendances implicationnelles.

La question pertinente qu'il est nécessaire de se poser avant de pouvoir exploiter ces données est celle de la possibilité de dégager de l'information diachronique à partir de structures synchroniques. Nous pensons que le modèle que nous avons développé au chapitre 6 est apte à tirer parti des informations synchroniques des bases de données précédentes pour évaluer et reproduire au moins approximativement des dynamiques d'évolution des systèmes phonologiques. L'idée est que le nombre de langues (et le choix qui a été fait pour la sélection de ces langues) permet d'estimer de façon correcte les contraintes internes (naturelles et distribuées). Sur la base de ce postulat, il est possible de dégager de la base les fréquences de chaque phonème en fonction des présences des autres segments dans les systèmes.

En se référant à l'annexe qui précise le calcul de la fonction d'énergie du modèle à partir des données précédentes, il est possible dès lors d'étudier les évolutions de systèmes phonologiques de la même façon que nous avons étudié des items abstraits au chapitre 6. Il sera intéressant alors de comparer les résultats des simulations avec des changements phonético-phonologiques attestés dans les langues du monde. Notons ici que s'il ne sera jamais possible de prédire une

évolution particulière, du fait des phénomènes stochastiques en jeu, il sera possible néanmoins de déterminer des spectres d'évolutions possibles à partir d'une configuration de segments initiale.

Notons pour conclure l'existence d'une autre base de données non-informatisée et de plus faible taille que nous pourrions mettre à profit. Il s'agit de la base présentée et analysée par Johanna Nichols dans son livre "*Linguistic Diversity in Space and Time*" [Nichols, 1992], qui a l'avantage de renseigner le lecteur sur les schémas de distribution spatiale d'un certain nombre de caractéristiques typologiques "classiques". Il est concevable d'appliquer le même raisonnement pour ces données typologiques plus générales que celui développé plus haut pour les segments des langues du monde. Il sera alors possible d'étudier en particulier l'apparition au cours des simulations de tendances évolutives, comme par exemple des tendances implicationnelles.

Application de la théorie des réseaux au modèle d'évolution des systèmes linguistiques

Si nous avons introduit dans nos travaux un modèle de réseau social pour abstraire les relations entre les membres d'une communauté, nous ne nous sommes appuyés, ni n'avons fait mention, des travaux plus généraux portant sur les réseaux.

Ceux-ci sont nombreux et variés, et se penchent sur des problèmes très divers, allant du routage d'information sur Internet ou des indexations croisées des pages Web (les hyper-liens) aux réseaux métaboliques connectés par des interactions chimiques, ou encore au réseau des publications scientifiques [Wang, 2002b] (p. 885). Les réseaux sont souvent étudiés et classifiés selon certaines variables calculées à partir de leur topologie : coefficient de "clusterisation" ("*cluster coefficient*"), longueur du chemin moyen ("*average path length*"), distribution de degré ("*degree distribution*") [ibid] (p. 887-889). Selon les valeurs de ces indicateurs, les réseaux possèdent différentes propriétés : robustesse ou fragilité plus ou moins importante, dynamiques particulières, synchronisation aisée ou impossible. . .

Une des voies de recherche est naturellement la découverte de réseaux offrant des propriétés intéressantes pour différents types d'applications. Différents schémas de connectivité particuliers ont été plus particulièrement étudiés : graphes aléatoires ("*random graphs*"), réseaux "*small-world*", réseaux "*scale-free*" ou réseaux évolutifs [ibid]. Certains de ces réseaux ont déjà été appliqués à des phénomènes linguistiques, en particulier pour décrire la structure formée des mots du lexique [Dorogovtsev and Mendes, 2001] [Cancho and Solé, 2001] [Sigman and Cecchi, 2002].

Il paraît profitable de profiter des acquis théoriques généraux sur les réseaux pour raffiner et approfondir notre étude de l'évolution des systèmes linguistiques. En effet, des phénomènes comme l'implémentation d'un changement dans une population peuvent être mis en relation avec les variables citées précédemment. La capacité d'un réseau à se synchroniser est ici particulièrement intéressante, puisqu'elle peut être rapprochée de la possibilité pour un changement de s'étendre à toute une communauté. Nous souhaitons nous appuyer sur les travaux de socio-linguistique pour déterminer quelle architecture de réseau est la plus proche des réseaux réels de locuteurs (existence d'individus centraux ou meneurs lors des phénomènes de changement, influence des individus marginaux. . .). A partir de notre mécanisme de convergence ou de divergence linguistique, nous pourrions alors évaluer les propriétés du réseau, et en particulier les régimes dynamiques et de synchronisation. Nous espérons que ceci pourra éclairer la question de la propagation d'un changement.

Il est à noter ici que dans les approches précédentes, tout comme dans des approches mathématiques formelles visant à établir la convergence d'une population d'individus vers une conven-

tion lexicale [Cucker et al., à paraître], les réseaux considérés sont toujours constitués d'arêtes de valeurs positives ou nulle. Nous avons cependant insisté sur le fait que les réseaux sociolinguistiques comprenaient selon nous des liens positifs et négatifs. Ceci représente une difficulté supplémentaire qu'il est intéressant d'investiguer.

Prise en compte des phénomènes d'acquisition dans le modèle d'évolution des systèmes linguistiques

Nous avons déjà proposé au chapitre 6 de nombreuses directions pour poursuivre l'étude du modèle préliminaire auquel nous avons eu recours pour étudier l'impact des contraintes sociales et internes.

Un élément que nous n'avons pas du tout abordé est celui de la prise en compte de l'acquisition. Il n'existe en effet pas de renouvellement des agents dans notre modèle, ni de phase d'apprentissage pour ceux-ci. Notre hypothèse implicite a été en fait qu'il est possible d'inclure l'acquisition dans les mécanismes d'évolution généraux auxquels nous avons eu recours.

Un raffinement important du modèle serait donc la prise en compte d'agents de durée de vie limitée, et un renouvellement de la population en accord avec l'organisation sociale de celle-ci.

Une idée simple est de reproduire ici les concepts du modèle de Nettle que nous avons souvent cités dans notre travail. Une division de la population en un certain nombre de classes d'âge, et la possibilité pour les plus jeunes d'acquérir un système linguistique en échantillonnant les agents autour d'eux (de façon modulée par les contraintes sociales) nous permettraient de mesurer l'impact de l'acquisition au niveau des différents concepts sur lesquels nous nous sommes appuyés lors de nos expériences : rythme d'évolution des agents ou du barycentre de ceux-ci, distance des premiers au dernier, entropie. . .

Parallèlement aux règles précédentes, la gestion de l'organisation sociale d'une communauté (persistance ou transformations) peut être basée sur des topologies particulières de réseaux, et des règles de connexions des nœuds empruntées par exemple au réseaux évolutifs mentionnés plus haut.

Corrélation entre schémas de diversité culturelle et schémas de diversité linguistique

Une idée que nous avons mentionnée à plusieurs reprises au cours de notre travail est la possibilité d'étudier les schémas de diversité de composantes extra-linguistique pour déterminer partiellement les réseaux sociaux entre groupes humains préhistoriques, et envisager ensuite les conséquences de leur organisation sur les évolutions linguistiques. Nous pensons ici en particulier au travail de Georges Sauvet et de ses collaborateurs sur la structuration des motifs des peintures pariétales en France et en Espagne au cours du Paléolithique Supérieur [Sauvet and Wlodarczyk, 1995].

Afin de développer notre compréhension et notre connaissance des liens sociaux qui unissaient les populations du passé, il paraît utile d'essayer de comparer les schémas de différentes activités culturelles. Il peut s'agir pour les périodes du Paléolithique de l'art pariétal, mais également des techniques de sculpture, des industries lithiques ou des objets de parure. Nous nous posons la question de savoir si des analyses assez fines peuvent être effectuées sur ces catégories d'artéfacts voisines, et si elles font ressortir des hétérogénéités qui reflètent une structuration sociale particulière de la macro-population.

Parallèlement à cette première approche, il paraît complémentaire d'étudier le lien qu'entretiennent les modes linguistiques avec les modes vestimentaires, de comportement etc., dans nos sociétés contemporaines. Certains chercheurs comme Labov abordent ce problème en cherchant à détecter des similitudes entre les acteurs impliqués dans les phénomènes précédents (en particulier par exemple les meneurs ("*leaders*") des changements)[Labov, 2001] (p. 361-363). Les questions peuvent être alors : peut-on détecter aujourd'hui des congruences nettes entre les différentes manifestations culturelles précédentes ? Existe-t-il des réseaux sociaux uniques permettant de rendre compte de façon unifiée de ces manifestations ?

En cas d'une réponse négative, l'étude des réseaux sociaux au cours de la préhistoire semble gravement handicapée. Il faut toutefois alors se demander si lors d'époques reculées, la structuration très particulière de la société n'induisait pas une convergence des attractions et des répulsions pour différentes manifestations culturelles au sein d'une macro-population. Il nous semble que cette question pourrait être partiellement investiguée à l'aide de modèles informatiques, qui permettraient de mieux mesurer les degrés de liberté accordés aux différents schémas de diversité culturelle au sein de différentes organisations sociétales.

7.2.2 Perspectives à plus long terme

Utilisation de l'outil informatique pour l'étude de l'évolution des langues

Si l'outil informatique est de plus en plus mis à profit pour étudier les conditions d'émergence de la fonction de langage, il reste encore utilisé de façon limitée pour l'étude des langues et de la variabilité linguistique. Comme nous l'avons déjà souligné à plusieurs reprises, les modélisateurs se penchent aujourd'hui plus sur les conditions d'émergence que sur les conditions d'évolution du langage, et il semble que beaucoup peut être fait pour mieux comprendre ces dernières.

Il ne s'agit pas ici de sous-estimer l'importante place qu'a prise l'informatique dans la linguistique. La linguistique computationnelle est ainsi un vaste champ d'étude, allant de la traduction automatique ou de la reconnaissance de parole à la construction et l'analyse de corpus (voir par exemple [Tsou et al., 1997]). Néanmoins, il nous semble qu'un créneau demeure relativement exploré entre les domaines précédents.

Nous pensons qu'il est ainsi intéressant de se pencher sur le champ de la linguistique historique, et de se poser la question de comment il est possible d'enrichir ce secteur grâce aux apports des machines. Nous pouvons rappeler ici qu'une des grandes découvertes dans cette discipline, à savoir la diffusion lexicale, s'est appuyée sur des bases de données informatiques, et ceci il y a plus de 30 ans [Wang and Chen, 1975].

Alors que de nombreux linguistes analysent et tentent de reconstruire des proto-langues à partir de langues actuelles, peu d'outils informatiques viennent aider les chercheurs dans ce travail minutieux ; on peut citer le moteur de reconstruction de Lowe et Mazaudon, mais il semble que cette tentative soit l'une des seules à avoir été proposées au cours des dernières années [Lowe and Mazaudon, 1994].

Nous sommes intéressés par la possibilité de produire des outils capables d'aider le linguiste dans sa tâche. Tout en étant conscient que le travail de reconstruction appelle un savoir qu'il est extrêmement difficile d'inculquer à une machine, nous pensons que des outils développés pour être des aides ou des assistants de travail pourraient rendre des services appréciables. A cette fin, il nous paraît judicieux non pas de bâtir des logiciels travaillant sur les données linguistiques comme sur tout autre type de données, mais d'insuffler au contraire un savoir linguistique, même

rudimentaire, dans des programmes d'analyse des correspondances lexicales. Ainsi, l'utilisation de décompositions en traits des segments, d'algorithmes de comparaison de formes phonétiques décrites à l'aide d'un alphabet phonétique... pourraient conduire à proposer au linguiste de possibles proximités entre les mots de listes de vocabulaire, et ainsi raccourcir son travail de recherche de correspondances, ou lui suggérer des éléments qu'il risquerait d'oublier. Afin de définir au mieux les caractéristiques qu'un tel produit devrait posséder, notre objectif est de nous appuyer sur une certaine analyse de l'activité d'un chercheur lors du reconstruction de reconstruction.

L'utilisation d'algorithmes issus de la cladistique pour la reconstruction des filiations à partir des pourcentages de cognats ou même de données vocaliques [Ben Hamed et al., 2002] est un autre sous-domaine où l'ordinateur peut se révéler très utile pour le test des innombrables arbres possibles de filiation. Des algorithmes ont également été élaborés pour tenter de détecter les emprunts non détectés par le linguiste au niveau des listes de cognats [Minett and Wang, 2002]. Cependant, ces approches de type algorithmique ne mettent pas réellement à profit la connaissance linguistique, mais raisonnent de façon abstraite sur des données vidées partiellement et temporairement de leur sens.

Autre exemple concret, nous pensons que le calcul de distances sémantiques ou phonétiques entre deux mots gagnerait à être investigué et surtout formalisé au moins partiellement. En effet, la linguistique historique se base pour ses reconstructions sur des proximités sémantiques et phonétiques plus ou moins forte (confère les critiques vis à vis des comparaisons multi-latérales). De telles distances peuvent être envisagées (i) au niveau phonétique, à partir des différences entre les segments, des possibles insertions ou délétions de segments dans les mots... (ii) au niveau sémantique, à l'aide par exemple de l'analyse des fréquences de co-occurrence de mots dans des gros corpus... En ce qui concerne le second point, nous pensons que d'autres travaux menés sur les relations de synonymie entre les mots d'une langue (voir par exemple [Victorri et al., à paraître] ou [Ploux and Victorri, 1998]) pourraient être incorporés dans des études sur l'évolution linguistique : de premières études sont ainsi menées pour comparer les cartes sémantiques de langues différentes [Ploux and Ji, à paraître], et il apparaît possible d'étendre les concepts en jeu à la comparaison de différentes périodes historiques d'une même langue (par exemple les dialectes chinois pour lesquels des dictionnaires assez anciens sont disponibles).

Différents modèles basés sur les idées précédentes nous paraissent judicieux pour l'évaluation des différentes méthodes de reconstruction, et pourraient donc être fort utiles dans le débat qui oppose les partisans d'une limite des reconstructions vers 8,000 BP, et ceux qui dépassent allègrement cette valeur seuil.

De l'objet au méta-objet ou comment apprendre des régularités fonctionnelles

Si nous nous sommes intéressés au mode d'émergence des stratégies linguistiques, nous n'avons pas du tout mentionné la façon dont ces stratégies pouvaient être manipulées ou transmises d'une génération à une autre. Le problème se trouve surtout au niveau des items et des stratégies syntaxiques.

Au niveau de l'émergence du langage, l'émergence de la syntaxe, à cause de la complexité de cette dernière, est sans aucun doute le secteur qui offre et offrira le plus de résistance aux efforts des modélisateurs (confère les travaux récents évoqués chapitre 3). Il est difficile d'éviter

le problème de la circularité des modèles (on observe en sortie ce que l'on a placé en entrée), et une véritable émergence de structures syntaxiques est un phénomène très difficile à mettre en place.

Les études portant sur le lexique (par exemple émergence de la convention dans une population d'agents) manient des objets mathématiques simples, qui peuvent entrer en compétition, ou être soumis à certaines contraintes, mais n'exerçant pas d'action sur les autres items autrement que par le biais de certaines dynamiques de compétition homogènes. . . . En tentant un parallèle avec les fonctions et les objets mathématiques constants, de tels objets peuvent être dits d'ordre 0 (ce sont en un sens des fonctions qui ne prennent aucune valeur d'entrée, autrement dit des fonctions constantes). En comparaison, les structures syntaxiques manipulent des items linguistiques en les composant. Elles peuvent être d'ordre 1 ou plus, et selon le cas, peuvent manipuler des objets d'ordre 0, voire des objets d'ordre supérieur. Elles jouent vis à vis des items d'ordre inférieur au leur le même rôle qu'une fonction mathématique (par exemple une fonction cosinus ou exponentielle) peut avoir sur un nombre réel ou entier.

L'apprentissage ou la création de structures syntaxiques par un système ou un mécanisme "pseudo-cognitif" est rendu très complexe par cette distinction d'ordre. Là où il est facile de créer ou d'apprendre des mots ou des sons, il est bien plus difficile de manipuler des structures fonctionnelles. Là où l'apprentissage des mots ou des sons repose sur la détection (la découverte) de régularités simples, l'apprentissage des structures syntaxiques repose lui sur la découverte de **régularités fonctionnelles**.

La détection de régularités fonctionnelles, ou dit autrement l'apprentissage de fonctions est un exercice très difficile. Cette question constitue une partie importante des théories de l'apprentissage. Une méthode connue repose sur l'utilisation de réseaux de neurones comme les perceptrons multi-couches, qui tentent d'approcher la fonction par le biais de données bruitées produites par celle-ci. Le problème n'est pas différent pour l'acquisition des structures linguistiques, mais les fonctions opèrent sur des unités discrètes, et leur conception de haut niveau les rend difficiles à capturer par un réseau de neurones. C'est bien sûr ce problème qui est au cœur de l'argument de la pauvreté du stimulus de Chomsky.

La théorie des principes et des paramètres, afin de détecter les régularités fonctionnelles linguistiques, recourt à un ensemble de principes fonctionnels qui font l'inventaire des fonctions qui peuvent être utilisées dans la langue. Comme pour de nombreux problèmes algorithmiques dont la vérification de la correction d'une solution est bien plus aisée que sa découverte, le problème de l'apprentissage de la syntaxe est rendu plus aisé par l'utilisation des principes et des paramètres (pour une introduction à ces problèmes, voir [Cormen et al., 1994], chapitre 36 sur la NP-complétude).

La question est donc posée s'il est possible de construire des mécanismes d'apprentissage généraux qui puissent détecter très efficacement les structures syntaxiques. Une méthode qui s'apparente superficiellement à l'existence de "principes" est la possibilité d'envisager la recherche des fonctions syntaxiques dans un espace de fonctions particulier. A la différence de la théorie des principes et des paramètres qui postule un caractère inné de ces éléments, la structure et les contraintes de l'espace fonctionnel de recherche pourraient être ancrées dans des capacités cognitives plus générales, en particulier au niveau sémantique [Schoenemann, 1999].

Un outil mathématique est particulièrement bien adapté au traitement des fonctions : il s'agit du lambda-calcul. Il s'inscrit avant tout dans une théorie générale des fonctions, et a été

développé dans les années 1940. Le grand avantage de ce paradigme est que les fonctions y sont “des objets comme les autres”, ce qui signifie qu’il n’existe pas de différenciation conceptuelle entre des objets d’ordres différents. Nous sommes particulièrement intéressés par la possibilité de définir des systèmes d’apprentissage et de manipulation des structures syntaxiques grâce à cet outil mathématique. Nous pensons plus précisément que la structure de l’espace de recherche de fonctions syntaxiques pourrait être transcrite à l’aide de conditions assez générales au niveau de règles de lambda-calcul, et que les possibilités de combinaison de ces règles permettraient de générer un grand nombre de structures syntaxiques plus ou moins complexes. Ces idées sont pour l’instant encore très préliminaires, mais nous espérons pouvoir y travailler dans un futur plus ou moins proche.

Intelligence artificielle et langage

Nous pouvons avant de conclure ce travail mentionner les liens fructueux qui gagnent à être tissés entre l’étude de l’origine du langage (ou plus généralement l’étude du langage) et l’intelligence artificielle dans son acceptation la plus générale.

Depuis une dizaine d’années, un flux de concepts et d’outils issus de l’intelligence artificielle (systèmes multi-agents, réseaux de neurones...) est venu irriguer le domaine de l’origine du langage. Ceci a permis et permettra pendant encore un long moment de produire des travaux originaux et porteurs d’idées nouvelles pour la compréhension de l’émergence de la communication humaine. Néanmoins, nous pensons qu’un mouvement de balancier peut s’instaurer, et l’étude du langage humain apporter en retour à l’intelligence artificielle : à l’heure où les limites des systèmes experts ou autres architectures prédéfinies sont de mieux en mieux comprises, les phénomènes émergents peuvent être des éléments-clés dans la génération de systèmes intelligents d’un nouveau genre, aux performances supérieures à celles des systèmes actuels.

Le langage est un des fondements de l’intelligence telle que nous la rencontrons chez notre espèce et telle que nous tentons de la reproduire dans des machines, et la même démarche émergente peut-être envisagée pour engendrer des systèmes linguistiques originaux. Nous pensons que comprendre et mimer la genèse du langage dans notre espèce est un des moyens de créer du sens linguistique dans une machine, même si la route est encore longue avant de parvenir à un système capable d’un langage réfléchi, même rudimentaire. Les pivots de cette recherche ambitieuse sont selon nous à trouver dans les capacités d’un système d’apprentissage à organiser de l’information linguistique et sémantique, et surtout à composer sans cesse celle-ci pour accroître la complexité des structures manipulées.

7.3 En guise de mot final...

”Il est bien plus beau de savoir quelque chose sur tout, que de savoir tout d’une chose”. Cette recherche esthétique de Pascal dans *les Pensées* nous semble se retrouver dans l’étude des problématiques scientifiques que nous avons tenté d’introduire dans ce travail. Beauté d’une construction scientifique vaste et exigeante, complexe et fragile, sans cesse remaniée et raffinée, qui se penche sur un élément fondateur de la culture et de la cognition humaines. L’étude de l’origine du langage offre aussi une occasion rare de combler un plaisir de l’appétit de connaissance, et de rêver par instant de “tout connaître sur tout”. Elle nous ouvre une porte sur ce quoi notre humanité s’est construite, sur ce qui fait l’homme et fonde la richesse de ses relations avec autrui. Ce qui nous semble être la définition même et surtout la valeur d’une science humaine et sociale.

知之為知之
不知為不知
是知也

Savoir qu'on sait quand on sait,
et savoir qu'on ne sait pas
quand on ne sait pas :
C'est là la vraie connaissance.

Entretiens de Confucius (II)