

L'auto-organisation dans l'évolution de la parole

Pierre-Yves Oudeyer
INRIA Bordeaux Sud-Ouest
<http://www.pyoudeyer.com>

Les systèmes de vocalisations humains, véhicules physiques du langage, sont caractérisés par des formes et des propriétés structurales complexes. Ils sont combinatoires, basés sur la ré-utilisation systématique de phonèmes, et l'ensemble des répertoires de phonèmes des langues du monde est marqué à la fois par de fortes régularités statistiques, les universaux, et une grande diversité. En outre, ce sont des codes culturellement partagés par chaque communauté de locuteurs. Quelle est l'origine des formes de la parole ? Quels sont les mécanismes qui, au cours de la phylogenèse et de l'évolution culturelle, ont permis leur évolution ? Comment un code de la parole partagé peut-il se former dans une communauté d'individus ? Je vais m'intéresser dans ce chapitre à la manière dont les phénomènes d'auto-organisation, et leurs interactions avec la sélection naturelle, peuvent permettre d'éclairer ces trois questions.

La tendance qu'ont de nombreux systèmes physiques complexes à générer spontanément des formes nouvelles et organisées, comme les cristaux de glaces ou les spirales galactiques, est présente en effet tout autant dans le monde inorganique que dans le monde vivant. Ainsi, l'explication de l'origine des formes du vivant ne peut reposer uniquement sur le principe de sélection naturelle, qui doit être complété par la compréhension des mécanismes de génération de formes nouvelles dans lesquels l'auto-organisation est centrale. Or, ceci s'applique aux formes sociales et culturelles du vivant, en particulier aux formes de la parole et du langage. Je vais donc ainsi commencer par articuler de manière générale les relations entre auto-organisation, sélection naturelle et néo-Darwinisme pour la compréhension de la genèse des formes du vivant. Je vais ensuite instancier ces relations dans le cadre des trois questions que j'ai énoncées ci-dessus. J'expliquerai alors pourquoi l'utilisation de simulations et de modèles informatiques est fondamentale pour faire progresser les théories qui y sont afférentes. Enfin, je présenterai un exemple d'expérimentation d'un modèle informatique qui montre que certains mécanismes simples de couplages sensorimoteurs permettent de générer des systèmes de parole combinatoires, caractérisés par la dualité universaux/diversité, et partagés culturellement. Je conclurai par les scénarios évolutionnaires que cette expérimentation informatique vient compléter ou renouveler.

Auto-organisation et évolution des formes du vivant

La physique, chaudron de formes auto-organisées

La nature regorge de formes et de motifs fascinants d'organisation, et en particulier dans sa partie inorganique. La silhouette des montagnes est la même que l'on regarde à l'échelle du rocher, du pic ou de la chaîne. Les dunes de sable s'alignent souvent en longues bandes parallèles. L'eau se cristallise en flocons symétriques et dentelés quand la température s'y prête. Et quand elle coule dans les rivières et tombe des cascades, apparaissent des tourbillons en forme de trompettes et les bulles se rassemblent en structure parfois polyédrales. Les éclairs dessinent dans le ciel des ramifications à l'allure végétale. L'alternance de gel et de dégel sur les sols pierreux de la toundra laisse des empreintes polygonales sur le sol. La liste de ces formes rivalise de complexité avec bien des artefacts humains, comme on peut l'apprécier sur la figure 1. Et pourtant rien ni personne ne les a dessinées ou conçues. Pas même la sélection naturelle, le concepteur aveugle de Dawkins¹. Quel est donc le mystère qui explique leur existence ?

¹ Dawkins, R. (1986). *The Blind Watchmaker*. New York: W. W. Norton & Company.



Figure 1 La nature regorge de formes et de motifs organisés sans qu'il y ait quelque part de plans qui aient servi à les construire : on dit qu'ils sont auto-organisés. Ici, des bandes parallèles qui courent sur les dunes, des bulles d'eau à la surface du liquide qu'on a agité, et les structures polyédrales qui restent quand elles sèchent, un cristal de glace, des montagnes dont les formes sont les mêmes qu'on les regarde à l'échelle du rocher ou à l'échelle du pic.

En fait, toutes ces structures organisées ont un point commun : elles sont le résultat macroscopique des interactions locales entre les nombreux composants du système dans lequel elles prennent forme. Leurs propriétés organisationnelles globales ne sont pas présentes au niveau local. En effet, la forme d'une molécule d'eau, ainsi que ses propriétés physico-chimiques individuelles n'ont rien à voir avec celles des cristaux de glace, des tourbillons, ou encore des polyèdres de bulles. Les empreintes polygonales de la toundra ne correspondent pas à la forme des pierres qui les composent, et ont une organisation spatiale très différente de l'organisation temporelle du gel et du dégel. Voilà la marque de l'auto-organisation. Les phénomènes auto-organisés dans la nature caractérisent des systèmes physiques très variés, mais certaines propriétés typiques peuvent être identifiées. Non-linéarité, brisures de symétrie et présence d'« attracteurs » sont ainsi souvent présents. Par exemple, lorsqu'on chauffe par le dessous une fine couche d'huile s'étalant sur une surface plane, des courants de convection prenant des formes géométriques particulières (lignes ou polygones) s'auto-organisent et ces formes changent brutalement quand on passe certains seuils de températures (voir figure 2). Entre ces seuils, au contraire, les formes restent globalement assez stables même si on les perturbe, constituant des attracteurs. Une autre propriété de nombreux systèmes auto-organisés est l'historicité, souvent associée à la sensibilité aux conditions initiales des systèmes chaotiques : l'attracteur dans lequel le système tombe, c'est-à-dire les formes qui sont produites par le système complexe, peut être très différent en fonction de petites variations des conditions initiales. C'est le cas par exemple de la magnétisation du fer : chacun des atomes d'une plaque de fer peut être vu comme une sorte d'aimant ayant plusieurs orientations possibles, qui est aléatoire si la température est élevée. Si elle passe en dessous d'un certain seuil, alors un phénomène d'auto-organisation se produit : tous les atomes adoptent spontanément la même orientation magnétique. Cependant, cette orientation commune est quasiment imprédictible au départ et de toutes petites variations aléatoires de l'orientation initiale de quelques atomes peut faire se magnétiser la plaque dans une direction très différente. Or ces variations des conditions initiales sont typiquement liées à des événements contingents ayant interagi avec la plaque de métal : c'est pourquoi l'état final de la plaque dépend de son histoire en plus de ses propriétés physiques intrinsèques, d'où le terme d'historicité.

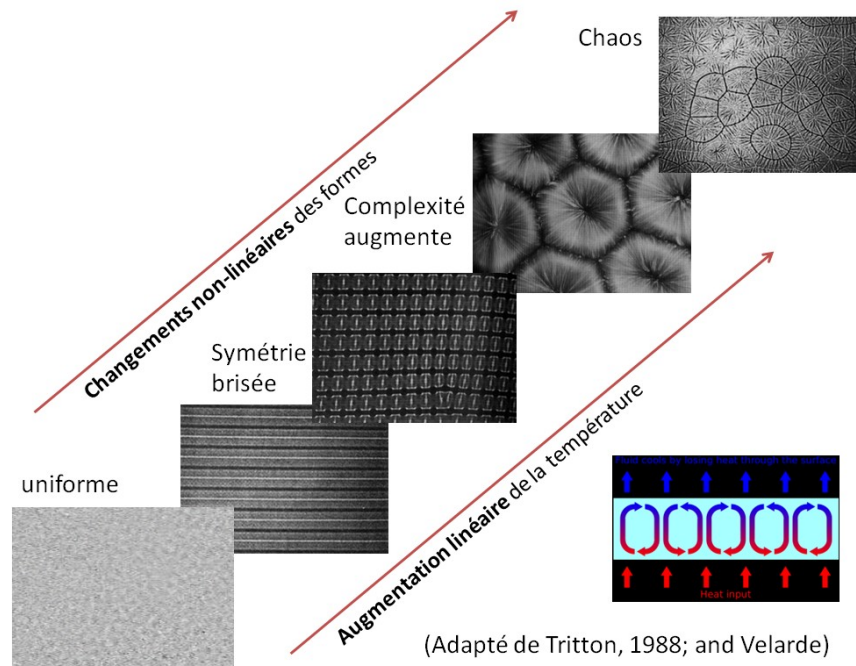


Figure 2 Cellules de Rayleigh-Bénard : si on chauffe par le dessous une fine couche d'huile s'étalant sur une surface plane, des courants de convections prenant des formes géométriques particulières (lignes, pavages hexagonaux) s'auto-organisent et ces formes changent brutalement quand on passe certains seuils de températures. Ce type de non-linéarité caractérise de nombreux systèmes auto-organisés du monde inorganique comme du monde vivant¹.

Le concept d'auto-organisation des systèmes complexes constitue la pierre de touche du changement paradigmatique que les sciences de la complexité ont opéré au 20ème siècle². Depuis Newton, la bonne science se devait être réductionniste, et consistait à étudier les systèmes naturels en les décomposant en sous-systèmes plus simples. Par exemple, pour comprendre comment le corps humain fonctionnait, on se devait d'étudier d'un côté le coeur, de l'autre le système nerveux, et d'un autre côté encore par exemple le système limbique. D'ailleurs, on ne s'arrêtait pas là, et l'étude du système nerveux par exemple se devait d'être divisée en l'étude du cortex, du thalamus ou des innervations motrices périphériques. Cette méthode nous a évidemment permis d'apprendre une somme impressionnante de connaissances. Mais les chantages de la complexité l'ont battue en brèche. Leur credo : « le tout est plus que la somme des parties ».

L'impact de l'auto-organisation sur l'origine des formes du vivant

Or les systèmes complexes, c'est-à-dire les systèmes composés de nombreux sous-systèmes en interaction, abondent dans la nature et ont la très forte tendance à s'auto-organiser. Les exemples de la partie précédente ont été volontairement choisis parmi les systèmes inorganiques pour montrer que la propriété d'auto-organisation peut caractériser des systèmes dont les mécanismes n'ont rien à voir avec celui de la sélection naturelle. Cependant, l'auto-organisation s'applique de la même manière aux systèmes vivants. C'est d'ailleurs un concept largement utilisé dans plusieurs champs de la biologie. Il est en particulier central aux théories qui expliquent les facultés des sociétés d'insectes à construire des nids ou des ruches, à chasser en groupe ou explorer de manière décentralisée et efficace les ressources

¹ Photos adapted from Tritton, D.J. (1988) Physical fluid dynamics, Oxford: Oxford University Press, and Manuel Velarde, Universidad Complutense, Madrid.

² Ashby, W.R. (1956) An Introduction to Cybernetics, Chapman & Hall ; Nicolis, G., Prigogine, I. (1977) Self-organization in non-equilibrium systems : from dissipative structures to order through fluctuations, New-York: Wiley ; Kauffman, S. (1996) At home in the universe : the search for laws of self-organization and complexity, Oxford : Oxford University Press ; Ball, P. (2001) The self-made tapestry : pattern formation in nature, Oxford : Oxford University Press.

de nourriture de l'environnement¹. En biologie du développement, il sert aussi par exemple à expliquer la formation des patterns colorés sur la peau des animaux comme les papillons, les zèbres, les jaguars ou les coccinelles².

Il semble donc qu'il soit possible qu'il y ait dans les systèmes biologiques des mécanismes créateurs de formes et patterns qui soient orthogonaux à la sélection naturelle, et ceci grâce à leur propriété d'auto-organisation. Or la sélection naturelle, et les explications fonctionnalistes qui lui sont associées, constitue le coeur de la majorité des argumentations en biologie quand il s'agit d'expliquer la présence d'une structure, d'une forme ou d'un pattern dans un organisme. Quel est donc l'articulation entre la théorie de la sélection naturelle et l'auto-organisation ?

Certains chercheurs ont proposé l'idée que l'auto-organisation remettait en question le rôle central de la sélection naturelle dans l'explication de l'évolution des organismes vivants. Waldrop³ explique:

''les systèmes dynamiques complexes peuvent parfois passer spontanément d'un état de désordre à un état d'ordre ; est-ce une force motrice de l'évolution ? ... Avons nous manqué quelque chose à propos de l'évolution - un principe clé qui a contrôlé le développement de la vie de manière différente de la sélection naturelle, des dérives génétiques, et de tous les autres mécanismes que les biologistes ont invoqué au cours des années ? ... Oui ! Et l'élément manquant ... est l'auto-organisation spontanée : la tendance qu'ont les systèmes dynamiques complexes à se placer dans des états ordonnés sans qu'il soit besoin d'aucune pression de sélection''.

Cependant, ce n'est pas la position qui est prise dans cet article. Plutôt que de voir l'auto-organisation comme un concept qui minimise le rôle de la sélection naturelle en proposant des mécanismes créateurs de formes concurrents, il est plus exact de le voir comme d'une part correspondant à un niveau d'explication différent et surtout d'autre part comme décrivant des mécanismes qui décuplent la puissance de la sélection naturelle. Les mécanismes ayant la propriété d'auto-organisation sont donc complètement intégrables au mécanisme de la sélection naturelle pour l'explication de l'évolution des formes du vivant.

L'approche classique du néo-Darwinisme

Pour le voir précisément, il faut d'abord rappeler en quoi consiste le mécanisme de la sélection naturelle selon le néo-darwinisme. Il caractérise un système composé d'individus ayant chacun des traits, formes ou structures particuliers. Ensuite, les individus de ce système sont capables de se répliquer. Cette réplification doit parfois générer des individus qui ne sont pas les exactes copies de leurs ancêtres, mais des petites variations. Ce sont ces variations qui sont à la source de la diversité des individus. Enfin, chaque individu a la capacité de se répliquer plus ou moins facilement selon sa structure et l'environnement qui l'entoure. Cette réplification différentielle des individus donne lieu à une « sélection » de ceux qui sont le plus aptes à se répliquer. La combinaison du processus de variation et du processus de sélection fait qu'au cours des générations, les structures ou traits des individus qui les aident à se reproduire sont conservés et améliorés.

Il est un point crucial sur lequel la théorie de la sélection naturelle reste neutre : c'est la manière dont sont générées les variations, et plus généralement la manière dont sont générés les individus, avec leurs formes, leurs traits et leurs structures. Un certain nombre d'arguments néo-darwiniens considèrent les mécanismes de variations des formes comme secondaires par rapport aux avantages reproductifs de ces formes quand il s'agit d'expliquer leur évolution. Cette mise au second plan implique implicitement que la relation entre le niveau des gènes, considérée comme l'espace principal dans lequel s'opèrent les variations par les mutations et les cross-over, et le niveau du phénotype,

¹ Camazine, S., Deneubourg, J-L., Franks, N.R., Sneyd, J., Theraulaz, G., and Bonabeau E. (2002) Self-Organization in Biological Systems, Princeton University Press.

² Ball, P. (2001) The self-made tapestry, Pattern formation in nature, Oxford University Press.

³ Waldrop, M. (1990) Spontaneous order, evolution, and life, Science, 247, pp. 1543—5.

considéré comme une image isomorphe de l'espace des gènes, est simple et linéaire. Selon cette vision, l'exploration de l'espace des formes (qui déterminent avec l'environnement le degré d'efficacité de réplication des gènes) peut être comprise simplement en regardant la manière dont on se déplace dans l'espace des génomes. Or les mécanismes de mutations qui permettent justement ces déplacements sont de petite amplitude (la plupart des mutations n'affectent qu'une toute petite partie des génomes quand il y a réplication), et donc des variations aléatoires des gènes permettent d'explorer uniformément tout l'espace des génomes. Ce qui veut dire que dans l'hypothèse où les deux espaces génotypiques et phénotypiques ont la même structure, alors l'espace des formes est exploré de manière quasi-continue, par petites modifications successives des formes pré-existantes. Heureusement pour l'apparition des formes de vie complexes, ce n'est pas le cas. En effet, si ce mécanisme de petites variations successives des formes est efficace pour le réglage fin des structures des organismes, il rendrait la recherche des formes aussi complexe que celle des organismes humains équivalente à la recherche d'une aiguille dans une botte de foin car les génomes sont de trop grande dimension (Keefe et Szostak, 2001).

L'auto-organisation contraint l'espace des formes à explorer : toutes les formes ne sont pas également faciles à faire émerger

C'est là que le concept d'auto-organisation vient à l'aide de ce mécanisme d'exploration naïf de l'espace des formes dans le cadre de la théorie néodarwinienne. En effet, la relation entre les gènes et les formes des organismes se caractérise par sa complexité et sa forte non-linéarité qui s'expriment pendant le développement ontogénétique et épigénétique de chaque organisme. En fait, les organismes sont construits à partir d'une cellule souche qui contient un génome, et cette cellule souche peut être vue comme un système dynamique paramétré par son génome et sous l'influence des perturbations imposées par l'environnement. Ce système dynamique constitue un système auto-organisé qui a le même type de propriétés que les systèmes auto-organisés présentés dans la partie précédente. Le génome est un ensemble de paramètres analogues à la température des liquides dans les systèmes de Bénard, et l'environnement est l'analogie du bruit pour la magnétisation du fer (mais c'est un bruit évidemment très structuré et structurant !). Ainsi, le développement d'un organisme à partir de sa cellule souche partage un certain nombre de propriétés avec la formation auto-organisée des systèmes physiques: des formes, des structures et des patterns apparaissent au niveau global, et sont qualitativement différentes de celles qui définissent le fonctionnement local, c'est-à-dire différentes de celles qui caractérisent la structure de la cellule souche et de son génome. Le pavage hexagonal qui peut apparaître à partir d'une simple différence de température dans un liquide homogène donne une idée de la manière dont une simple suite de nucléotides entourée d'un système moléculaire qui les transforme automatiquement en protéines peut générer un organisme bipède doté de deux yeux pour voir, d'oreilles, d'un cerveau immensément complexe.

En outre, comme les systèmes de Bénard ou les plaques ferromagnétiques, les systèmes dynamiques définis par les génomes et les cellules qui les contiennent sont caractérisés par un paysage d'attracteurs : il y a de larges zones de l'espace des paramètres pour lesquelles systématiquement le système dynamique adopte un comportement dont la structure reste à peu près la même. Pour les systèmes de Bénard, il y a une plage de température qui permet de générer des bandes parallèles et qui est assez large pour qu'on puisse la trouver facilement. Pour les plaques ferromagnétiques, la zone de température pendant laquelle le système arrive à trouver une cohérence magnétique globale est aussi très large. Ainsi, pour les organismes vivants, non seulement il est possible de générer des structures auto-organisées aux propriétés globales complexes, mais en plus ces structures sont certainement générées par les génomes appartenant à de larges sous-espaces dans l'espace des génomes, qu'on appelle des bassins d'attraction. La structuration de l'espace en bassins d'attraction de ce genre de système dynamique permet donc ainsi de faire que l'exploration de l'espace des formes soit facilitée et ne s'apparente pas à la recherche d'une aiguille dans une botte de foin.

Comme dans les systèmes ferromagnétiques, le bruit (structuré) imposé par l'environnement sur le développement du système dynamique peut conduire à ce qu'il prenne des voies de développement différentes. Pour les morceaux de fer à basse température, cela correspondait à la magnétisation dans

un sens ou dans l'autre. Pour un organisme vivant, cela correspond à ses formes : c'est ainsi qu'il arrive que même des vrais jumeaux peuvent avoir des différences morphologiques assez importantes. Ceci illustre aussi pourquoi la relation entre les gènes et les formes des organismes ne se caractérise pas seulement par sa complexité et sa non-linéarité, mais aussi par son non-déterminisme. Comme dans les systèmes de Bénard où l'exploration de l'espace du paramètre de température peut parfois conduire à des changements rapides et conséquents du comportement du système (par exemple le passage des bandes parallèles aux cellules carrées), l'exploration de l'espace des génomes peut aussi conduire parfois à des changements de formes rapides et conséquents. Cela correspond possiblement à de nombreuses observations de changements de formes très rapides dans l'évolution, comme en témoignent les fossiles qu'étudient les anthropologues, et qui sont à la base de la théorie des équilibres ponctués proposée par (Eldredge et Gould, 1972).

Pour résumer, les propriétés d'auto-organisation du système dynamique composé par les cellules et leurs génomes apportent une structuration cruciale à l'espace des formes en le contraignant, ce qui rend la découverte de formes complexes et robustes beaucoup plus facile pour la sélection naturelle. D'une part elles permettent à un génome de générer des formes complexes et très organisées sans qu'il soit besoin d'en spécifier précisément chaque détail dans le génome (de la même manière que les formes polygonales de Bénard ne sont pas spécifiées précisément, ou sous la forme d'un plan, dans les propriétés des molécules de liquide). D'autre part, elle organisent le paysage de ces formes possibles en bassins d'attraction à l'intérieur desquelles elles se ressemblent beaucoup (c'est là que se font les évolutions graduelles, avec des réglages fins des structures existantes), et entre lesquelles les formes peuvent différer substantiellement (c'est le passage de l'un à l'autre qui peut provoquer des inventions soudaines et puissantes dans l'évolution). Pour donner une image simple, l'auto-organisation fournit un catalogue de formes complexes réparties dans un paysage de vallées dans lesquelles et entre lesquelles la sélection naturelle se déplace et fait son choix : l'auto-organisation propose, la sélection naturelle dispose. Evidemment ceci n'est qu'une image, car la sélection naturelle par son déplacement permet justement de faire apparaître de nouveaux mécanismes eux-mêmes auto-organisés, qui structurent l'espace des formes dans lequel elle se déplace ; donc la sélection naturelle participe à la formation de ces mécanismes qui l'aident à avoir un déplacement efficace dans l'espace des formes ; vice versa, le mécanisme de la sélection naturelle est certainement apparu dans l'histoire de la vie grâce aux comportements auto-organisés de systèmes qui étaient encore complètement étrangers à la sélection naturelle ; la sélection naturelle et les mécanismes auto-organisés s'aident donc mutuellement dans une sorte de spirale qui permet à la complexité d'augmenter au cours de l'évolution.

La conséquence de cette intrication entre sélection naturelle et auto-organisation est que l'explication de l'origine et de l'évolution des formes du vivant nécessite au moins deux types d'argumentaires. Le premier, classique, est l'argumentaire fonctionnaliste néo-darwiniste : il consiste à identifier le contexte écologique dans lequel un trait nouveau a pu apparaître et d'en articuler le bilan confrontant les coûts et les avantages reproductifs. Le second argumentaire est plus rarement utilisé mais est tout aussi essentiel : il s'agit d'identifier les mécanismes développementaux/épigénétiques, et les contraintes qui y sont associées, qui ont pu faciliter, ou au contraire rendre difficile, la genèse de ces traits nouveaux. Or la notion d'auto-organisation est centrale dans la manière dont ces mécanismes développementaux impactent la genèse des formes.

Auto-organisation et évolution des formes du langage et des langues

La question de savoir comment parole et langage sont venus à l'être humain, et la question de savoir comment des langues nouvelles se forment et évoluent, sont parmi les plus difficiles qui soient posées à la science. Alors qu'elles ont été écartées de la scène scientifique pendant la presque totalité du 20^{ème} siècle, à la suite de la déclaration de la Société Linguistique de Paris qui la bannit de sa constitution, elle sont redevenues le centre des recherches de toute une communauté de scientifiques. Un consensus se dégage de la communauté des chercheurs qui aujourd'hui s'y attèlent : la recherche doit être multidisciplinaire. En effet, c'est un puzzle aux ramifications immenses qui dépassent les compétences de chaque domaine de recherche pris indépendamment. C'est d'abord parce que les deux grandes questions, celle de l'origine du langage et celle de l'origine et de l'évolution des langues, doivent être

décomposées en sous-questions elles-mêmes déjà fort complexes : Qu'est-ce que le langage ? Qu'est-ce qu'une langue ? Comment s'articulent entre eux les sons, les mots, les phrases, les représentations sémantiques ? Comment le cerveau représente-t-il et manipule-t-il ces sons, ces phrases, et les concepts qu'elles véhiculent ? Comment apprend-on à parler ? Peut-on discerner l'inné de l'acquis ? A quoi sert le langage ? Quel est son rôle social ? Comment une langue se forme-t-elle et change au cours des générations successives de ses locuteurs ? Que sait-on de l'histoire de chaque langue ? Pourquoi le langage et les langues sont tels qu'ils sont ? Pourquoi y-a-t-il des tendances universelles et en même temps une grande diversité des structures linguistiques ? Quelle est l'influence du langage sur la perception et la conception du monde ? Que sait-on de l'histoire de la capacité de parler chez les humains ? Est-ce plutôt le résultat d'une évolution génétique, comme l'apparition des yeux, ou une invention culturelle, comme l'écriture ? Est-ce une adaptation à un environnement changeant ? Une modification interne de l'individu qui a permis d'augmenter ses chances de reproduction ? Est-ce une exaptation, effet collatéral de changements qui n'étaient pas initialement reliés au comportement de communication ? Quels sont les pré-requis évolutifs qui ont permis l'apparition de la capacité de parler ? Comment eux-mêmes sont-ils apparus ? Indépendamment ? Génétiquement ? Culturellement ?

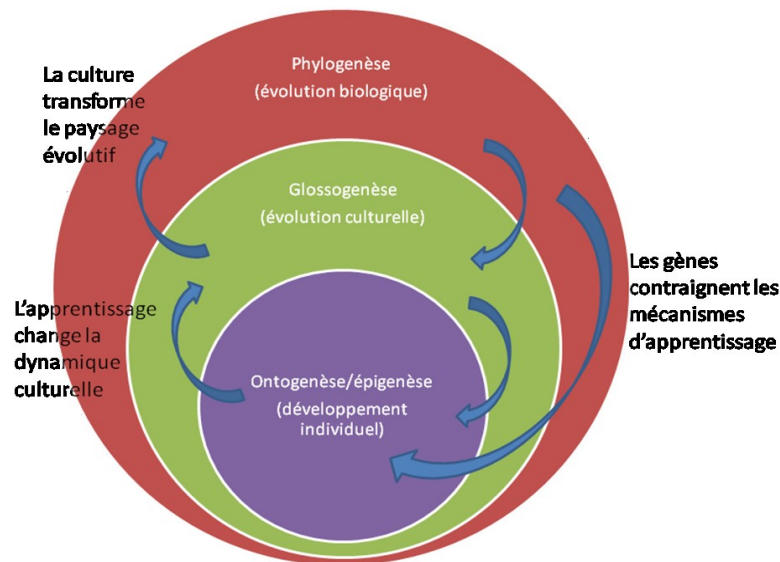


Figure 3 Les multiples échelles d'interactions à l'origine du langage

Face à la diversité de ces questions se dresse une diversité encore plus grande de disciplines et de méthodes. Les linguistes, même s'ils continuent à fournir des données cruciales sur l'histoire des langues ainsi que sur les tendances universelles de leurs structures, ne sont plus les seuls acteurs. La psychologie développementale, la psychologie cognitive et la neuropsychologie font des études comportementales de l'acquisition du langage ainsi que des troubles du langage, souvent révélateurs des mécanismes cognitifs qui sont impliqués dans le traitement du langage. Les neurosciences, en particulier avec les dispositifs d'imagerie cérébrale qui permettent de visualiser quelles zones du cerveau sont actives quand on effectue une tâche donnée, essaient de trouver les corrélats neuronaux des comportements de parole, pour en découvrir l'organisation cérébrale. Des chercheurs étudient aussi la physiologie de l'appareil vocal, pour essayer de comprendre la manière dont nous produisons des sons. La physiologie de l'oreille, capteur essentiel dans la chaîne de décodage de la parole (c'est la vision quand il s'agit de langue des signes), est aussi au centre des recherches. Les archéologues examinent les fossiles et les artefacts qui sont les restes des premiers hommes, et tentent d'une part d'en déduire l'évolution morphologique des hommes (en particulier de son larynx), et d'autre part de se faire une idée des activités qu'ils pratiquaient (quels outils fabriquaient-ils ? et comment les utilisaient-ils ? comment ces outils peuvent-ils nous renseigner sur le degré de développement cognitif ?). Les anthropologues vont à la rencontre des peuples isolés, et rendent compte des différences culturelles, en particulier celles liées aux langues et aux conceptions qu'elles véhiculent. Les primatologues essaient de rendre compte des capacités de communication de nos ancêtres les chimpanzés et de les comparer

aux nôtres. D'une part, les généticiens séquencent les génomes de l'homme et des espèces qui sont ses ancêtres potentiels quand c'est possible pour préciser leurs liens phylogénétiques, et d'autre part utilisent les informations génétiques des différents peuples de la planète pour aider à la reconstruction de l'histoire des langues, qui est souvent corrélée avec l'histoire des gènes des locuteurs qui les parlent.

Le langage implique donc une multitude de composantes qui interagissent de manière complexe sur plusieurs échelles de temps en parallèle : l'échelle ontogénétique, qui caractérise le développement de l'individu, l'échelle glosso-génétique ou culturelle, qui caractérise l'évolution des cultures, et l'échelle génétique, qui caractérise l'évolution des espèces (voir figure 3). En particulier, le langage met en œuvre des interactions complexes, à la fois physiques et fonctionnelles, entre de multiples circuits cérébraux, plusieurs organes, les individus qui en sont équipés et l'environnement dans lequel ils vivent. Or, comme on l'a vu dans la section précédente, s'il est fondamental d'étudier chacune de ces composantes indépendamment, afin de réduire la complexité du problème, il est aussi nécessaire d'en étudier les interactions. Un certain nombre de chercheurs ont en effet proposé l'idée selon laquelle de nombreuses propriétés du langage et des langues ne sont codées dans aucun de leurs composants, c'est-à-dire ni dans certaines structures cérébrales spécifiques, ni dans les propriétés des appareils auditifs ou vocaux, ou ni même chez un individu considéré indépendamment, mais pourraient être des résultats auto-organisés de l'interaction complexe et dynamique de ces composants et de ses individus¹. Or ces phénomènes d'auto-organisation sont souvent compliqués à comprendre, à prévoir intuitivement, et à formuler verbalement, d'où le recours de plus en plus important à la modélisation mathématique et informatique comme nous allons le voir.

Modèles et simulations informatiques de l'évolution du langage

Expérimenter les systèmes complexes

L'une des manières les plus efficaces aujourd'hui pour développer notre compréhension de la dynamique des systèmes auto-organisés est l'utilisation des ordinateurs ou des robots. En effet, ils permettent d'élaborer des modèles opérationnels dont on connaît toutes les hypothèses, de les faire fonctionner, et d'en observer le comportement selon les valeurs des paramètres fixés dans le cadre des modèles. C'est pourquoi, en plus des linguistes, des psychologues, des anthropologues, des chercheurs en neurosciences, des généticiens et des physiologistes, les mathématiciens et les informaticiens/roboticiens ont désormais un rôle crucial dans cette recherche.

Un modèle opérationnel est un système qui définit formellement l'ensemble de ses présuppositions et surtout qui permet de calculer ses conséquences, c'est-à-dire de prouver qu'il mène à un ensemble de conclusions données. Il existe deux grands types de modèles opérationnels. Le premier, celui utilisé par les mathématiciens et certains biologistes théoriciens, consiste à abstraire du phénomène du langage un certain nombre de variables et leurs lois d'évolution sous la forme d'équations mathématiques. Cela ressemble le plus souvent à des systèmes d'équations différentielles couplées, et bénéficie du cadre de la théorie des systèmes dynamiques. Le second type, qui permet d'étudier certains phénomènes complexes qui se prêtent difficilement à la modélisation mathématique, est celui utilisé par les chercheurs en informatique: il consiste à construire des systèmes artificiels implantés dans des ordinateurs ou sur des robots. Ces systèmes artificiels sont composés de programmes qui le plus souvent prennent la forme d'agents artificiels dotés de cerveaux et de corps artificiels; on pourra les appeler robots même s'ils évoluent dans des environnements virtuels. Ces robots sont alors mis en interaction dans un environnement artificiel ou réel, et on peut étudier leur dynamique. C'est ce qu'on appelle la « méthode de l'artificiel ». L'utilisation de machines computationnelles pour simuler et étudier les phénomènes naturels n'est d'ailleurs pas nouvelle: Lorenz a utilisé les premiers ordinateurs

¹ Hurford, J., Studdert-Kennedy, M., Knight, C. (1998) *Approaches to the Evolution of Language : Social and Cognitive Bases*. Cambridge : Cambridge University Press.

Lindblom, B., MacNeilage, P., and Studdert-Kennedy, M. (1984) Self-organizing processes and the explanation of language universals. In Butterworth, B. and Bernard, C. and Dahl, O., editors, *Explanations for Language Universals*, pages 181--203.

Oudeyer, P.-Y. (2006) *Self-organization in the evolution of speech*, Oxford University Press.

pour étudier le comportement de modèles climatologiques, Fermi pour simuler l'interaction entre des particules magnétisées, Turing pour imaginer comment les processus de morphogenèse pouvaient s'auto-organiser, Von Neumann pour étudier l'auto-réplication.

Plus récemment et grâce à cette méthode, l'éthologie a fait un bon en avant pour la compréhension des comportements et des performances des insectes sociaux¹. Des simulations informatiques de sociétés d'insectes, basées sur le concept d'agents informatiques modélisant chaque insecte individuellement, ce qu'on appelle parfois modèles individuocentrés, ont été construites. Cela a permis d'établir des caractéristiques suffisantes du comportement et des capacités des insectes pour observer la formation de structures collectives, comme la construction des nids chez les termites, la formation de collectifs de chasse ou de recherche de nourriture chez les fourmis, la formation des bancs de poissons, la thermorégulation dans les ruches des abeilles ou la formation de structures sociales chez les guêpes. De manière générale, ces simulations informatiques ont montré qu'il n'était souvent pas nécessaire que les insectes soient équipés de structures cognitives complexes pour que, collectivement, ils forment des structures complexes. Ces modèles informatiques sont même parvenus à faire des prédictions qui ont été vérifiées par la suite sur le terrain.

Les physiciens utilisent aussi de plus en plus l'ordinateur pour construire des simulations de systèmes complexes qui leurs permettent de développer leurs intuitions. En manipulant des automates cellulaires - sortes de grilles dont les cases peuvent être allumées ou éteintes et dont l'évolution dépend de l'état de leurs voisines selon des règles simples - ils ont découvert comment, à partir de structures soit complètement aléatoires soit complètement ordonnées, des motifs complexes avec des symétries non-triviales pouvaient se former. Les exemples sont fort divers: les cristaux de glace, les distributions des avalanches dans les tas de sables ou dans les montagnes, les dunes dans le désert, les formes des deltas fluviaux, la formation des galaxies ou celle des polyèdres de bulles au pied des cascades. Pour les physiciens, les automates cellulaires ne sont évidemment pas à proprement parler des modèles physiques des cristaux de glace ou des avalanches, mais ils ont joué un rôle de métaphore et d'analogie qui a déclenché un renouvellement de la manière dont ils percevaient ces phénomènes.

Informatique et origines du langage et des langues

Il est également possible d'utiliser les ordinateurs et les simulations à base d'agents non seulement pour nous aider à comprendre les phénomènes qui caractérisent l'auto-organisation de la matière, des structures biologiques simples, ou des sociétés d'insectes, mais aussi pour l'étude des phénomènes qui caractérisent l'homme et ses sociétés. Le temps est venu de faire entrer l'ordinateur et les robots parmi les outils des sciences humaines. Ainsi, la construction de systèmes artificiels dans le cadre de la recherche sur les origines du langage et de l'évolution des langues bénéficie d'une popularité grandissante dans la communauté scientifique en tant qu'outil pour étudier les phénomènes du langage liés à l'interaction complexe de ses composants².

Il y a deux grands types d'utilisation de ces systèmes: 1) ils servent à évaluer la cohérence interne des théories verbales déjà proposées en clarifiant toutes les hypothèses et en vérifiant qu'elles mènent bien aux conclusions proposées (et, souvent, on découvre des failles dans les présupposés ainsi que dans les conclusions qui doivent être révisées) ; 2) ils servent à engendrer de nouvelles théories ou à explorer celles qui, souvent, apparaissent d'elles-mêmes quand on essaie tout simplement de construire un système artificiel qui reproduit les comportements de parole des humains.

Un certain nombre de résultats décisifs ont déjà été obtenus et ont permis d'ouvrir la voie à la résolution de questions jusque là sans réponses: la génération décentralisée de conventions lexicales et

¹ Bonabeau, E., Theraulaz, G., Deneubourg, J.L., Aron, S. & Camazine, S. 1997. Self-organization in social insects. Trends in Ecology and Evolution, 12: 188-193.

² Steels, L. (1997) The synthetic modeling of language origins, Evolution of Communication, 1(1), p. 1—35.

sémantiques dans des communautés de robots¹, la formation de répertoires partagés de voyelles ou de syllabes dans des sociétés d'agents, avec des propriétés de régularités structurelles qui ressemblent beaucoup à celles des langues humaines², la formation de structures syntaxiques conventionnalisées³ ou les conditions dans lesquelles la compositionnalité peut être sélectionnée⁴.

Il est important de noter que, dans le cadre de la recherche sur les origines du langage, cette méthodologie de l'artificiel est avant tout une *méthodologie exploratoire*. Elle s'insère dans une logique scientifique d'abduction, c'est-à-dire une logique dans laquelle on cherche des prémisses qui peuvent mener à une conclusion donnée (au contraire de la déduction dans laquelle on cherche les conclusions auxquelles peuvent mener des prémisses données).

Le mot "modèle" a ici un sens différent de son acception traditionnelle. Selon cette dernière, modéliser consiste à observer un phénomène naturel puis à essayer d'en abstraire les mécanismes et les variables fondamentales pour construire à partir d'elles un formalisme capable de prédire précisément la réalité. Dans le cas qui nous intéresse, il s'agit plutôt de s'interroger qualitativement sur les types de mécanismes que la nature a pu mettre en œuvre pour résoudre tel ou tel problème. En effet, le langage est un phénomène tellement complexe que la simple observation ne permet pas de *déduire* des mécanismes explicatifs. Au contraire, il est nécessaire d'avoir au préalable une bonne conceptualisation de l'espace des mécanismes et des hypothèses qui pourraient expliquer les phénomènes complexes du langage. Et c'est là le rôle des systèmes artificiels, ceux qu'on appelle parfois "modèles": développer notre intuition sur les dynamiques de formation du langage et des langues, et ébaucher l'espace des hypothèses.

Il ne s'agit donc pas d'établir la liste des mécanismes responsables de l'origine de tel ou tel aspect du langage. L'objectif est plus modestement d'essayer de faire une liste des candidats possibles, de contraindre l'espace des hypothèses, en particulier en montrant des exemples de mécanismes qui sont suffisants et des exemples de mécanismes qui ne sont pas nécessaires.

Le code de la parole

Je vais maintenant illustrer ce travail de modélisation informatique de l'évolution du langage et des langues par la description d'une expérimentation qui se focalise sur le problème de l'origine de la parole, c'est-à-dire les systèmes de sons en tant que véhicules et supports physiques du langage (au même titre que peuvent l'être les gestes dans les langues des signes). L'objectif de cette expérimentation est de participer à la reconceptualisation de ce problème en explicitant, en évaluant et en ouvrant plusieurs hypothèses scientifiques.

Digitalité et combinatorialité. Les humains ont un système de vocalisations complexe. Celles-ci sont digitales et combinatoriales, c'est-à-dire qu'elles sont construites à partir d'unités élémentaires, « sculptées » dans un continuum auditif et vocal, et systématiquement recombinaisonnées, puis ré-utilisées dans les vocalisations. Ces unités sont présentes à plusieurs niveaux (e.g. les primitives motrices d'obstruction du flux de l'air dans le conduit vocal, qu'on appelle gestes ; les coordinations de gestes, que l'on appelle phonèmes et dont font partie les consonnes et les voyelles ; les syllabes). Alors que l'espace articulatoire est continu et permet potentiellement une infinité de gestes et de phonèmes,

¹ Steels, L.; The synthetic modeling of language origins in *Evolution of Communication 1* (1997). Kaplan, F.; La naissance d'une langue chez les robots; Hermès, 2001.

2 de Boer, B.; The origins of vowel systems, *Oxford Linguistics*, Oxford University Press, 2001. Oudeyer, P.-Y.; Origins and learnability of syllable systems, a cultural evolutionary model, in P. Collet, C. Fonlupt, J. H. E. L. M. S. (Ed.), *Artificial Evolution*; LNCS 2310 (2001); pp. 143–155.

3 Batali J.; Computational simulations of the emergence of grammar; in Hurford, J. R., Studdert-Kennedy, M. and Knight C., editors, *Approaches to the Evolution of Language: Social and Cognitive Bases*; Cambridge University Press; 1998.

4 Kirby, S.; Spontaneous evolution of linguistic structure - an iterated learning model of the emergence of regularity and irregularity; *IEEE Transactions on Evolutionary Computation* 5 (2); 2001; 102–110.

chaque langue discrétise cet espace à sa manière en utilisant un répertoire de gestes et de phonèmes à la fois petit et fini¹. C'est pourquoi on appelle aussi parfois cette propriété le *codage phonémique*.

Universaux et diversité. En outre, malgré la grande diversité de ces unités dans les langues du monde, on y rencontre en même temps de fortes régularités. Par exemple, certains systèmes de voyelles sont beaucoup plus fréquents que d'autres, comme le système à cinq voyelles *e, i, o, a, u*. Il en va de même pour les consonnes. La manière dont les unités sont combinées est aussi très particulière: d'une part toutes les séquences de phonèmes ne sont pas autorisées dans une langue donnée, d'autre part l'ensemble des combinaisons de phonèmes est organisé en types génériques. Cette organisation en types génériques veut dire qu'on peut, par exemple, résumer les combinaisons de phonèmes autorisées en Japonais pour former des syllabes ("moras" plus exactement) par les types "CV/CVC/VC", où par exemple "CV" est un type qui désigne les syllabes composées de deux emplacements, avec dans le premier emplacement uniquement des phonèmes de la catégorie que l'on appelle "consonnes", alors que dans le second emplacement seuls les phonèmes de la catégorie "voyelles" sont autorisés.

Partage culturel. En outre, il faut remarquer que la parole est un code conventionnel. Alors qu'il y a des régularités statistiques au travers des langues humaines, chaque communauté linguistique possède sa propre manière de catégoriser les sons, et son propre répertoire de règles de combinaisons de ces sons. Par exemple, les Japonais n'entendent pas la différence entre le *r* de *read* et le *l* de *lead* en anglais. Comment alors une communauté linguistique en arrive-t-elle à former un code qui est partagé par tous ses membres, sans qu'il n'y ait de contrôle supervisé global ?

Depuis les travaux de de Boer ou de Kaplan², on sait comment un nouveau son ou un nouveau mot peut se propager et être accepté dans une population donnée. Mais ces mécanismes de négociation, encore appelés "dynamiques du consensus", font appel à la pré-existence de conventions et d'interactions linguistiques. Les modèles associés concernent donc plutôt la formation et l'évolution des langues, mais ne proposent pas de solution relatives à l'origine du langage. En effet, quand il n'y avait pas déjà de systèmes de communication conventionnels, comment sont apparues les premières conventions de la parole ?

C'est à cette dernière question en particulier que le modèle que je vais présenter s'intéresse. Elle est évidemment liée à celle de la formation des langues, car il s'agit de comprendre comment un code de la parole a pu être formé pour constituer la base des toutes premières langues. La différence principale entre les deux questions réside dans les propriétés qui doivent caractériser le mécanisme que l'on cherche. Pour la question de l'origine de la parole, on doit en particulier chercher un mécanisme explicatif qui ne présuppose ni l'existence de conventions linguistiques, ni l'existence de structures cognitives spécifiques au langage. Cela impliquerait en effet qu'on a affaire à des individus qui parlent déjà, et donc pour lesquels le langage est déjà apparu.

Auto-organisation et évolution de la parole

Il est donc naturel de se demander d'où vient cette organisation de la parole et comment un tel code conventionnel et partagé a pu se former dans une société d'agents qui ne disposaient pas déjà de conventions linguistiques. Comme je l'ai argumenté précédemment, deux types de réponses doivent être apportés. Le premier type est une réponse fonctionnelle : il établit la fonction des systèmes sonores, et montre que les systèmes qui ont l'organisation que nous avons décrite sont efficaces pour remplir cette fonction. Cela a par exemple été proposé par Liljencrantz et Lindblöm³ qui a montré que

¹ Studdert-Kennedy, M. and Goldstein, L. (2003) Launching language : the gestural origin of discrete infinity, in Christiansen, M. and Kirby, S. (eds.), *Language evolution : the state of the art*, pp. 235—254, Oxford :Oxford University Press.

² De Boer (2001) *The origins of vowel systems*, Oxford University Press.

Kaplan, F. (2001) *La naissance d'une langue chez les robots*, Hermès.

³ Liljencrantz, J.; Lindblom, B. (1972) Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language* 48: 839-862.

les régularités statistiques des répertoires de phonèmes peuvent être prédites en recherchant les systèmes de vocalisations les plus efficaces. Ce type de réponse est nécessaire, mais non suffisant : il ne permet pas d'expliquer *comment* l'évolution (génétique ou culturelle) pourrait avoir trouvé cette structure quasi-optimale, ni *comment* une communauté linguistique fait le "choix" d'une solution particulière parmi les nombreuses solutions quasi-optimales. En particulier, il se peut que la recherche darwinienne "naïve" avec des mutations aléatoires ne soit pas suffisamment efficace pour trouver des structures complexes comme celles de la parole : l'espace de recherche est trop grand.

C'est pourquoi un second type de réponse est nécessaire : il faut aussi trouver le moyen d'établir comment la sélection naturelle a pu « trouver » ces structures. On peut pour cela montrer que l'auto-organisation est susceptible, dans ce cas précis, de contraindre l'espace de recherche et d'aider la sélection naturelle. Il suffit de montrer qu'un système beaucoup plus simple que la structure que l'on cherche à expliquer s'auto-organise spontanément *en générant cette structure*.

Nous allons donc présenter maintenant un tel système et montrer comment des prémisses relativement simples d'un point de vue évolutionnaire peuvent conduire à la formation auto-organisée de codes de la parole.

Une expérimentation informatique de la formation des structures fondamentales de la parole

Ce modèle informatique est individu-centré : il consiste en la mise au point de robots virtuels caractérisés par un modèle des appareils auditifs et moteurs et par des réseaux de neurones artificiels qui connectent la modalité perceptuelle et la modalité motrice. Ces réseaux de neurones artificiels déterminent leur comportement, qui consiste principalement à effectuer du babillage vocal. Ce babillage, couplé avec les propriétés de plasticité des réseaux neuronaux, permet à ces robots d'apprendre les correspondances entre l'espace des perceptions auditives et l'espace des mouvements du conduit vocal. Enfin, ces robots sont placés ensemble dans un même environnement dans lequel ils peuvent non-seulement entendre leur propre babillage, mais également celui de leurs voisins, qui les influence. Nous allons ainsi montrer que des propriétés émergentes caractérisant les vocalisations produites par les robots d'une même population se forment spontanément.

Plus techniquement¹, les agents disposent d'une oreille artificielle (dont les propriétés peuvent être modifiées pour étudier leur rôle spécifique, voir ci-dessous), capable de transformer un signal acoustique en impulsions nerveuses qui stimulent les neurones d'une carte de neurones artificiels perceptuels. Ils disposent aussi d'une carte de neurones moteurs dont l'activation produit des mouvements d'un modèle du conduit vocal, qui lui même produit une onde acoustique (et dont le degré de réalisme peut également être modifié). Les cartes nerveuses (perceptuelle et motrice) sont totalement connectées entre elles (voir la figure 4). Au départ, les paramètres internes de tous les neurones, ainsi que les connections entre les deux cartes, sont aléatoires. Pour produire une vocalisation, un robot active aléatoirement plusieurs neurones de sa carte motrice, dont les paramètres internes codent pour des configurations articulatoires à atteindre en séquence, ce qui produit une trajectoire articulatoire et, par le biais du modèle du conduit vocal, une trajectoire acoustique qui est perçue grâce au modèle de l'appareil auditif. Ceci définit leur comportement de babillage. C'est pourquoi, au départ de l'expérience, les robots produisent des vocalisations qui sont aléatoires dans l'espace vocal. Cependant, ces réseaux de neurones sont caractérisés par deux formes de plasticité : 1) les connections intermodales évoluent de telle manière que le robot apprend les correspondances entre les trajectoires auditives perçues et les commandes motrices qui les génèrent quand il babille²; 2) Les neurones de chacune des cartes évoluent de telle manière qu'ils tendent à modéliser la distribution des

¹ Nous ne donnons ici qu'une description générale du système : pour une description mathématique précise, voir Oudeyer (2006) *Self-organization in the evolution of speech*, Oxford University Press.

² Les connections entre les deux cartes de neurones évoluent selon la loi de Hebb : celles qui relient des neurones qui sont souvent activés en même temps deviennent plus fortes, et celles qui relient des neurones dont l'activité n'est pas corrélée deviennent plus faibles. Ces connections sont aléatoires au début, et grâce au babillage des agents, elles s'organisent de telle manière que l'agent devient capable de trouver les commandes motrices correspondant à un son qu'il "entend".

sons que le robot entend¹ : les connections entre les deux cartes perceptuelles sont aussi telles que la distribution des sons codés par la carte perceptuelle (celle des sons perçus) reste à peu près la même que la distribution des sons codés par la carte motrice (celle des sons produits). Autrement dit, l'architecture nerveuse de l'agent est telle qu'il a tendance à produire la même distribution de sons que celle qu'il entend. Ainsi, si l'on fait écouter à un robot un flux de parole d'une langue donnée, son babillage va se régler/s'aligner sur la distribution des sons présents dans cette langue. Par exemple, si cette langue contient les voyelles [a, e, i] mais pas [o], le robot va se mettre à babiller en prononçant beaucoup plus souvent des [a, e, i] que des [o]. Ce comportement correspond à ce qui est observé chez les jeunes enfants, et parfois référé sous le terme de « phonological attunement »².

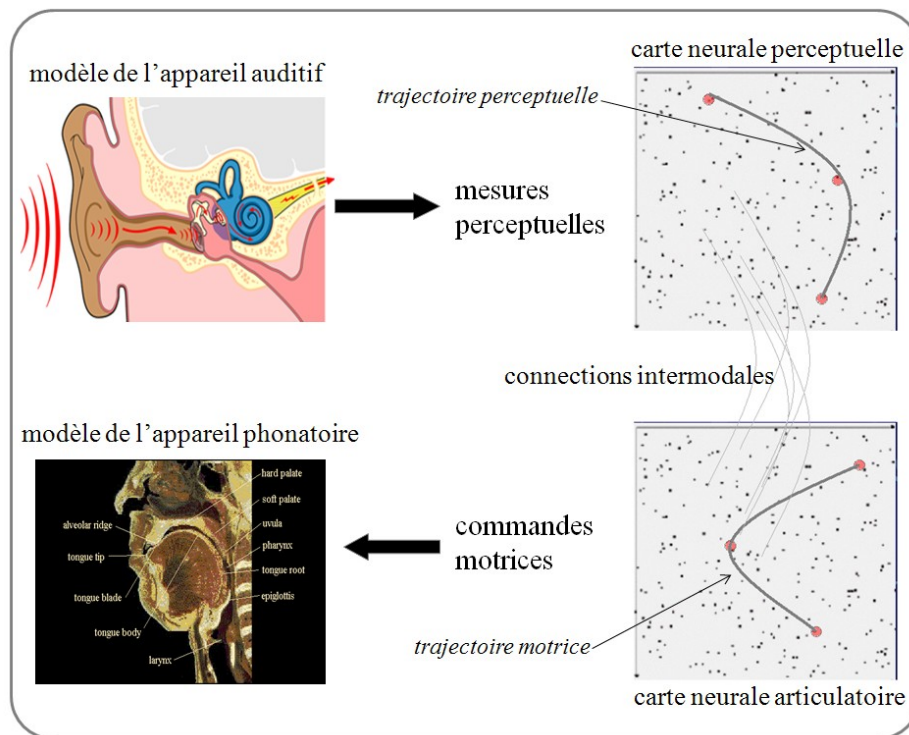


Figure 4: L'architecture d'un robot dans le système artificiel.

Chaque robot est doté d'une oreille artificielle, d'un conduit vocal artificiel, et de deux réseaux de neurones artificiels, l'un perceptuel et l'autre moteur, qui couplent ces deux organes. Ces réseaux de neurones sont au départ aléatoires mais caractérisés par deux formes de plasticité : 1) les connexions intermodales évoluent de telle manière que le robot apprend les correspondances entre les trajectoires auditives perçues et les commandes motrices qui les génèrent quand il babille ; 2) Les neurones de chacune des cartes évoluent de telle manière qu'ils tendent à modéliser la distribution des sons que les robots entendent.

Ainsi, si l'on fait écouter à un robot un flux de parole d'une langue donnée, son babillage va se régler/s'aligner sur la distribution des sons présents dans cette langue.

¹ Les neurones s'adaptent aux stimuli par sensibilisation : leur dynamique est telle que si un stimulus S est perçu, alors ils sont modifiés de telle manière que si l'on présente le même stimulus S juste après ils répondront encore plus.

² Vihman, M. (1996) Phonological development : the origins of language in the child, Cambridge, Mass. : Blackwell.

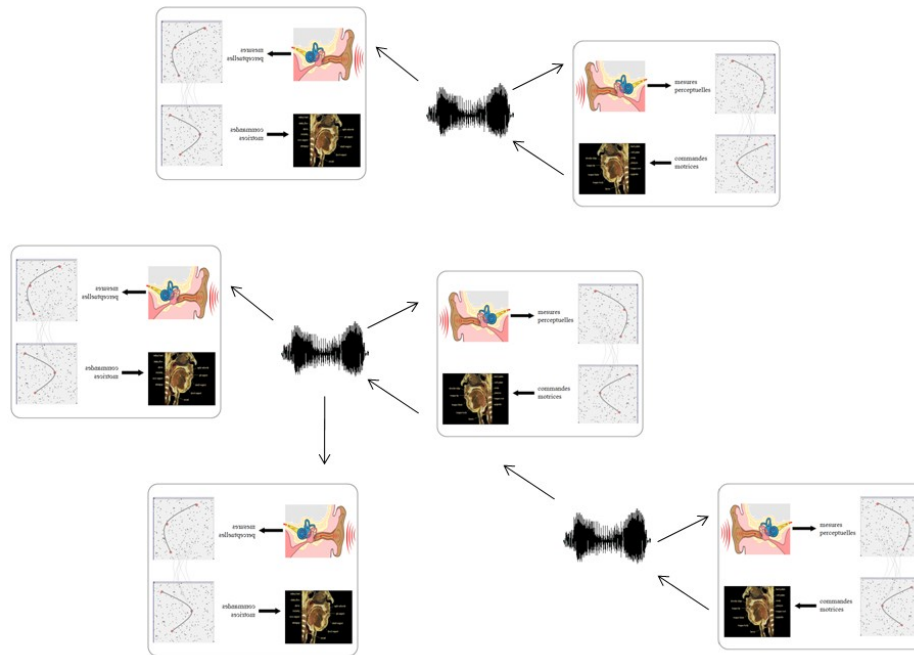


Figure 5 L'expérience consiste à placer dans un même environnement un ensemble de robots qui babillent, et donc entendent les vocalisations de leurs voisins en plus des leurs. Leurs réseaux neuronaux sont initialement aléatoires, et donc leurs vocalisations sont au départ également aléatoires et inorganisées. Etant donné que les propriétés de plasticité de leurs cerveaux les font babiller en s'alignant sur les vocalisations qu'ils entendent autour d'eux, et comme statistiquement ils produisent tous des vocalisations aléatoires dans l'espace vocal, l'état initial est un état d'équilibre. Cependant, cet équilibre est instable et le bruit inhérent à leurs interactions et à la dynamique de leurs neurones provoque une brisure de symétrie : leurs vocalisations se cristallisent et deviennent digitales, combinatoriales et partagées par tous les membres de la population (voir figure 6).

Un mécanisme unifié pour l'auto-organisation de la combinatorialité, de la dualité universaux/diversité, et du partage culturel

Ce type d'architecture a été utilisé fréquemment dans la littérature pour modéliser les phénomènes d'acquisition de la parole chez les enfants¹, dans des expériences dans lesquelles le système apprenait à prononcer les sons/syllabes d'une langue qu'on lui faisait entendre. Cependant, l'expérience qui est ici présentée est différente : on ne suppose pas qu'il existe au départ un système de parole déjà constitué. Au contraire, on va placer une population de robots babilleurs dans un environnement commun, de telle manière qu'ils vont percevoir à la fois leurs propres babillages et ceux de leurs voisins (voir figure 5). Etant donné que les propriétés de plasticité de leurs cerveaux les font babiller en s'alignant sur les vocalisations qu'ils entendent autour d'eux, et comme statistiquement ils produisent tous des vocalisations aléatoires dans l'espace vocal, l'état initial est un état d'équilibre.

Seulement, si l'on fait fonctionner la simulation, on s'aperçoit que cet équilibre n'est pas stable. En effet, il y a du bruit - de la "stochasticité" - qui fait que par hasard et de temps en temps, certains types de vocalisations seront prononcés plus souvent que d'autres. Or, le mécanisme de couplage décrit plus haut introduit une boucle de rétroaction positive: ces déviations de la moyenne sont amplifiées lorsqu'elles sont assez grandes, et la symétrie du système se brise. Les cartes de neurones s'auto-organisent alors en groupes concentrés de neurones, codant pour des configurations acoustiques et articulaires très précises dans l'espace des vocalisations (voir figure 6). En bref, l'espace continu des vocalisations a été discrétisé. Les vocalisations que les agents produisent ne sont plus holistiques, mais

¹ Kohonen, T. (1988) The neural phonetic typewriter, *Computer*, 21(3), pp. 11—22.

Sanguineti, V. and Laboisière, R. and Ostry, D. J. (1998) A Dynamic Biomechanical Model for Neural Control of Speech Production. *Journal of the Acoustical Society of America*, 103(3):1615-1627.

digitales: elles sont systématiquement construites par la mise en séquence de quelques configurations clés, que l'on peut alors appeler phonèmes. On voit apparaître le codage phonémique/digital et combinatorial qu'on a décrit plus haut. En outre, le « code phonémique » qui apparaît est le même chez tous les agents d'une même simulation, alors qu'il est différent d'une simulation à l'autre. On observe donc la formation d'une "convention culturelle", qui peut être diverse d'un groupe à l'autre.

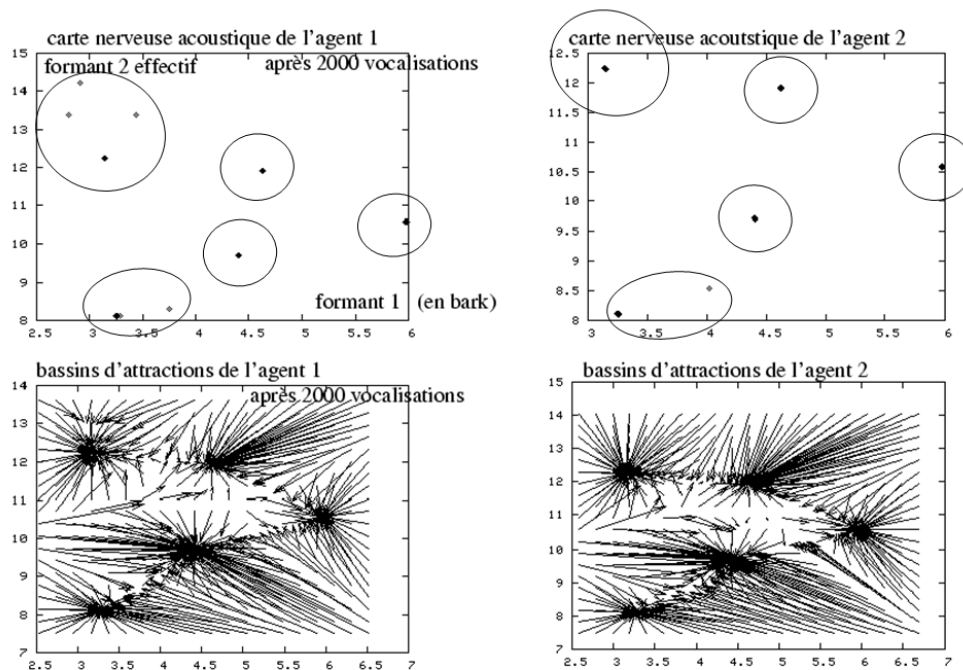


Figure 6 Très rapidement, la symétrie initiale des cartes neurales des robots se brise, et les neurones qui encodent initialement des configurations vocales aléatoirement réparties dans l'espace vocal encodent maintenant un petit nombre de configurations qui sont systématiquement ré-utilisées par les agents quand ils babillent : l'espace vocal a été discrétisé. En outre, ces configurations élémentaires qui émergent sont les mêmes chez tous les agents d'une même population, mais différentes entre populations. On peut le voir sur cette figure qui représente la carte des neurones perceptuels de deux agents après 2000 vocalisations (en haut), ainsi qu'une représentation de leur distribution (en bas). L'espace auditif est ici projeté sur le premier et le second formant effectif, exprimés en barks, ce qui permet de représenter les voyelles du système sonore auto-organisé dans cette simulation.

En fait, plusieurs variantes de cette expérience peuvent être réalisées et permettent d'affiner ces conclusions. Tout d'abord, il est possible de réaliser cette expérience avec un seul agent qui s'écoute babiller. Dans ce cas, on observe également un phénomène de cristallisation de ses vocalisations : très vite il ne produit plus que des trajectoires vocales qui transitent toutes par quelques configurations articulaires clés qui sont systématiquement ré-utilisées. On peut donc en déduire que l'apparition du codage phonémique, c'est-à-dire de la digitalité/combinatorialité, n'est pas le résultat des interactions sociales, mais des propriétés du couplage interne à chaque robot entre les modalités perceptuelles et motrices de la parole. Cependant, alors que les vocalisations de robots babillants isolés vont se cristalliser sur des systèmes de vocalisations différents, ces systèmes de vocalisations se synchronisent spontanément se les robots partagent le même environnement et sont capables de s'entendre les uns les autres : dans ce cas ces systèmes sont tous quasiment les mêmes dans une même population.

Une seconde variante importante de cette expérience consiste à faire varier les propriétés morpho-physiologiques des modèles des appareils auditifs et phonatoires afin de déterminer quel est l'impact de ces propriétés sur les systèmes qui se forment (ou qui ne se forment pas). En particulier, une propriété importante de ces appareils est la non-linéarité de la fonction qui fait correspondre une onde acoustique et une perception auditive à des commandes motrices et des configurations du conduit vocal. Le conduit vocal humain est en effet tel que pour certaines configurations articulaires, de

petites variations produisent de petites variations du son produit et perçu. Pour d'autres configurations, de petites variations produisent de grandes variations du son produit. Or cette propriété est centrale dans plusieurs théories qui proposent d'expliquer le codage phonémique de la parole, par exemple dans la théorie quantale de Stevens¹, ou dans le modèle DRM². Or il est possible d'utiliser un modèle des appareils acoustiques et phonatoires qui soit réaliste et incluse ces types de non-linéarités, mais il est aussi possible de construire un modèle non-réaliste volontairement linéaire afin de déterminer l'impact des non-linéarités. Ces expériences ont donc été menées. Tout d'abord, avec un modèle linéaire, on observe dans une population de robots babillants la cristallisation décrite précédemment : leurs vocalisations s'auto-organisent en un système combinatoire dans lequel quelques configurations articulatoires sont systématiquement ré-utilisées comme points clés des trajectoires vocales. Nous pouvons donc faire une première conclusion : ces simulations montrent que le codage phonémique peut apparaître spontanément sans linéarité dans les appareils phonatoires et auditifs. Cela n'implique pas que ces non-linéarités n'accélèrent pas la formation du codage phonémique, mais qu'elles ne sont pas nécessaires, comme proposé par la théorie quantale ou le modèle DRM.

Si, quand on utilise le modèle audio-phonatoire linéaire, on fait de nombreuses simulations et que l'on s'intéresse à la distribution de ces configurations articulatoires clés (qu'on peut voir comme des sortes de phonèmes), on s'aperçoit qu'elles sont globalement positionnées à des endroits aléatoires et uniformément dans l'espace des configurations possibles. Cependant, quand on utilise un modèle réaliste, reproduisant en particulier les propriétés de production et de perception des voyelles des humains³, on observe un phénomène supplémentaire. Outre le phénomène de cristallisation qui est le même qu'avec le modèle linéaire, les systèmes de vocalisations qui se forment sont caractérisés par des régularités statistiques qui ressemblent beaucoup aux systèmes humains. On peut par exemple faire des statistiques sur les voyelles qui apparaissent en tant que configurations articulatoires clés dans les systèmes émergents de cette expérience en faisant de nombreuses simulations. Les résultats, illustrés sur la figure 7, montrent que d'une part une certaine diversité de systèmes de voyelle apparaît, mais en même temps certains systèmes de voyelles apparaissent beaucoup plus fréquemment que d'autres. Il y a donc la même dualité universaux/diversité que celle que l'on observe dans les langues humaines, et cette simulation en propose une explication unifiée :

- 1) le système dynamique constitué par l'ensemble des agents babillants et par les couplages sensorimoteurs internes qui les caractérisent comporte un certain nombre d'attracteurs que sont les systèmes de vocalisation combinatoires/avec un codage phonémique partagés par tous les membres d'une population;
- 2) sous l'influence du bruit et des petites variations des conditions initiales, le système dynamique tombe dans un attracteur particulier, ce qui permet d'expliquer le « choix collectif décentralisé » de la population pour adopter un système plutôt qu'un autre ;
- 3) les non-linéarités des appareils auditifs et phonatoires introduisent des asymétries entre les attracteurs : certains d'entre eux ont un bassin d'attraction plus grand, en particulier ceux dans lesquels les phonèmes sont dans des zones où de petites variations articulatoires provoquent de petites variations perceptuelles, ce qui a pour conséquence d'augmenter la probabilité que le système y « tombe ».

En outre, non seulement il y a une correspondance structurelle entre les simulations et la réalité, mais les systèmes de voyelles apparaissant le plus fréquemment dans les populations de robot sont à peu près les mêmes, et dans les mêmes proportions, que les systèmes les plus fréquents dans les langues du monde. Il y a donc aussi une certaine correspondance quantitative. On peut donc en conclure que les non-linéarités des appareils auditifs et phonatoires sont ici déterminantes pour expliquer pourquoi certains systèmes de phonèmes sont statistiquement plus fréquents que d'autres. Par contre, encore une fois, l'existence même des phonèmes, c'est-à-dire l'existence d'un système de vocalisations dans

¹ Stevens, K.N., 1989. On the quantal nature of speech. *Journal of Phonetics*, 17, 3–45.

² Mrayati, M., Carre, R., Guerin, B. (1988) Distinctive regions and modes : a new theory of speech production, *Speech Communication*, 7, pp. 257–286.

³ Voir Oudeyer (2006) Ibid pour une description précise du modèle basé sur les travaux de de Boer (2001) Ibid.

lequel des configurations articulatoires ou auditives invariantes sont systématiquement réutilisées, ne nécessite pas ces non-linéarités.

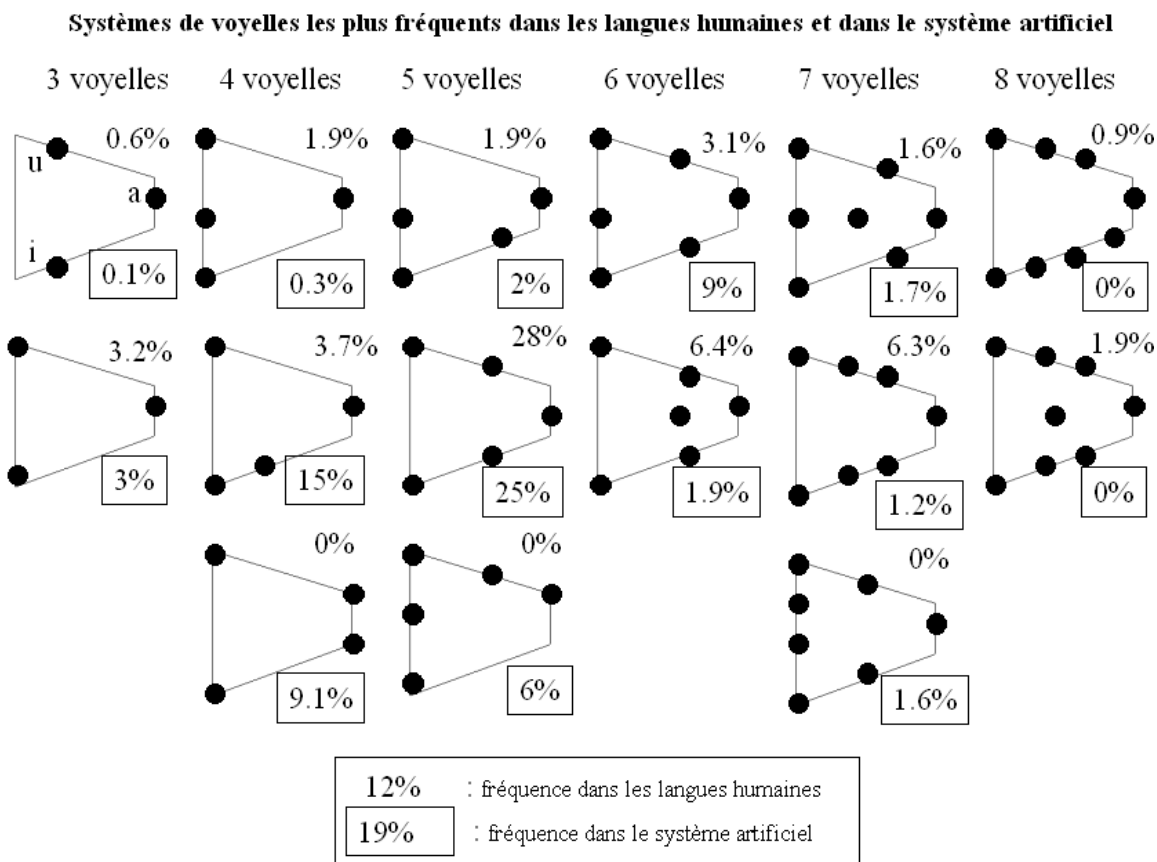


Figure 4 Comparaison entre la distribution des systèmes de voyelles auto-organisés dans le système artificiel et celle des systèmes de voyelles dans les langues humaines (d'après la base de donnée UPSID¹ (Maddieson, 1984). Les systèmes de voyelles sont représentés sur le triangle vocalique, dont la dimension horizontale correspond au premier formant, et la dimension verticale au second formant effectif. On observe que les systèmes les plus fréquents engendrés par les agents artificiels sont aussi les plus fréquents chez les humains, en particulier le système à 5 voyelles symétrique /a,e,i,o,u/ avec 25 pourcent dans les systèmes artificiels et 28 pourcent dans les langues humaines.

Il est cependant important de noter que dans ces simulations, les architectures neurales sont caractérisées par plusieurs paramètres, et que toutes les valeurs de ces paramètres ne permettent pas d'obtenir ces résultats. Néanmoins, il se trouve qu'un seul de ces paramètres a une influence importante sur les résultats² : les neurones sont caractérisés par une sélectivité σ aux stimuli plus ou moins grande. Si cette sélectivité est trop petite, aucune cristallisation ne se passe, bien que les robots restent capables d'apprendre les relations entre l'espace auditif et l'espace moteur. Si elle est trop grande, alors le système se cristallise en un système dégénéré dans lequel toutes les vocalisations sont exactement les mêmes et sont stationnaires : il n'y a qu'un phonème. Cependant, il y a une très grande plage de valeurs entre ces deux extrêmes qui permet de faire apparaître une cristallisation dans laquelle un système combinatoire avec de multiples phonèmes se forme dans la population de robot babillants.

Vers une vision nouvelle des scénarios évolutifs des origines de la parole

Ces remarques ont une importante conséquence si l'on utilise ce modèle pour imaginer des scénarios évolutifs ayant pu mener à la formation des premiers systèmes de vocalisations partageant des

¹ Maddieson, I. (1984) Patterns of sounds, Cambridge : Cambridge University Press.

² Voir Oudeyer (2006)

propriétés fondamentales des systèmes de parole de l'humain contemporain. En effet, elles impliquent d'abord que beaucoup de variations des paramètres de cette architecture neurale permettent de faire apparaître des systèmes de paroles combinatoriaux partagés par toute une population. Ensuite, elles impliquent qu'avec une telle architecture neurale, un tel système de parole peut apparaître sans supposer d'autres propriétés particulières des appareils auditifs et phonatoires que celle de pouvoir babiller initialement avec une certaine variété de sons. En particulier, aucune non-linéarité n'est nécessaire. Enfin, l'architecture neurale elle-même est relativement simple : elle met en œuvre des unités neurales dont les propriétés intrinsèques et les propriétés de plasticité sont très classiques et correspondent, au niveau fonctionnel, à la manière dont fonctionnent la plupart des unités neurales des cerveaux des mammifères¹. La spécificité de cette architecture est le fait que les cartes neurales auditives sont fortement et directement connectées aux cartes motrices, et que ces connections sont plastiques. Or, cette architecture et ces connections apparaissent comme étant les éléments essentiels au départ non pas de la parole, mais tout simplement de la capacité à apprendre à imiter des sons, ce que nous appelons ici l'imitation vocale adaptative². Ce qui nous amène au scénario évolutionnaire suivant pour conceptualiser l'origine des systèmes de vocalisations phonémiques et partagés par les membres d'une communauté :

- 1) L'imitation vocale adaptative est présente chez de nombreux animaux³ qui disposent de systèmes de vocalisations partagés et appris, mais chez qui le langage n'existe pas. Les éthologues ont d'ailleurs établis de nombreux avantages reproductifs potentiels caractérisant la capacité d'imitation vocale adaptative dans une communauté d'individus (comme par exemple le fait que cela permet de marquer l'appartenance au groupe). Il est donc raisonnable de penser qu'avant de savoir parler, les humains ont pu disposer de la capacité à s'imiter vocalement et que cette capacité d'imitation soit apparue avant le langage ;
- 2) Etre capable d'imitation vocale adaptative, ainsi que la plupart des avantages reproductifs identifiés par les éthologues pour expliquer la présence de cette capacité chez certains animaux, n'implique pas et ne nécessite pas un système de vocalisation combinatorial et codé phonémiquement. La zone de paramètres pour lesquels le paramètre de sélectivité σ est petit permet d'ailleurs aux robots d'apprendre très bien les correspondances perceptuo-motrices vocales sans pour autant générer de système phonémique ;
- 3) Si maintenant l'on se place dans un contexte écologique dans lequel la présence d'un système de parole combinatorial apporterait un avantage reproductif à ses possesseurs, alors les expériences que nous avons décrites permettent de dire qu'un simple changement de la valeur du paramètre de sélectivité σ des cartes de neurones auditives et motrices permettrait de faire apparaître spontanément des systèmes de vocalisations qui partagent plusieurs des propriétés fondamentales des systèmes de parole humains contemporains, au point même de pouvoir approximativement en prédire les systèmes de voyelles quand on utilise un modèle de la production et de la perception des voyelles chez l'humain. Cela permet alors de comprendre que ce qui fut un grand pas dans l'origine du langage, c'est-à-dire la formation de systèmes de vocalisations combinatoriales, a pu être la conséquence d'un petit changement biologique grâce aux propriétés auto-organisatrices de la matière neurale et de ses couplages multi-modaux.

Ce scénario dans lequel les systèmes de parole phonémiques auraient été sélectionnés grâce aux avantages reproductifs qu'ils auraient pu procurer, sélection rendue possible par la relative facilité de générer ces systèmes à partir des bases biologiques de l'imitation vocale adaptative, n'est cependant pas le seul que ce modèle peut appuyer. En effet, j'ai expliqué plus haut que dans la plage de valeurs du paramètre σ qui permet à des systèmes phonémiques de se former, la capacité d'imitation vocale est intacte et aussi performante. Or, à performances d'imitation égales, le passage du paramètre σ entre cette zone de paramètre et la zone de petite sélectivité ne présente pas de coût métabolique a priori.

¹ Voir Oudeyer (2006) Ibid

² Et qui correspond à la terminologie anglaise « adaptive vocal mimicry ».

³ Snowdown, C.T., Hausberger, M. (1997) Social influences on vocal development, Cambridge University Press.
Hauser, M. D. (1997) The Evolution of Communication. MIT Press/BradfordBooks.

Cela implique que dans un contexte écologique dans lequel ces structures neurales sont apparues sous l'effet d'une pression sélective pour l'imitation vocale adaptative, des mutations/variations neutres ont pu se produire et générer spontanément des systèmes de parole phonémique sans pression de sélection linguistique. Une observation rend ce scénario particulièrement stimulant : parmi les espèces animales capables d'apprendre à faire des imitations vocales, et pour lesquels il existe des systèmes de vocalisations partagés culturellement mais qui n'ont pas de langage, un certain nombre produisent des vocalisations ou des chants structurés autour d'unités de base systématiquement ré-utilisées. C'est le cas par exemple des canaris ou des diamants mandarins chez les oiseaux¹, ou des baleines à bosses². La fonction de cette structuration quasi-phonémique est encore très mal conceptualisée en éthologie. Or, le modèle que j'ai présenté, parce qu'il est neutre vis-à-vis de beaucoup de propriétés des appareils auditifs et vocaux, et parce que l'architecture neurale qu'il suppose correspond à l'équipement minimal pour l'imitation vocale adaptative, peut s'appliquer également à la formation des chants chez ces animaux. Dans ce cas, il fournit une hypothèse que l'incertitude sur la fonction du codage combinatorial de ces chants vient renforcer : des unités vocales combinatorialement ré-utilisées pourrait s'être formé spontanément comme un effet collatéral de l'équipement biologique pour l'imitation vocale adaptative. Il est donc également raisonnable d'imaginer que cela aurait pu se produire chez l'être humain : les systèmes de parole combinatoriaux auraient été recrutés seulement par la suite pour réaliser leur fonction linguistique. Dans ce cas, plusieurs des propriétés fondamentales des systèmes de parole de l'humain contemporain seraient des exaptations³.

Conclusion

Grâce à la construction et à l'utilisation d'un modèle informatique, j'ai montré comment une architecture de couplage sensorimoteur audio-phonatoire relativement simple permettait, dans une dynamique auto-organisatrice, la formation spontanée de systèmes de vocalisations combinatoriaux codés phonémiquement, partagés par tous les individus d'une même communauté, et caractérisés par la dualité universaux/diversité. La première contribution de ce travail est qu'il permet pour la première fois de proposer une explication unifiée de ces trois phénomènes.

En outre, cette architecture de couplage multimodal correspond au matériel neuronal minimal nécessaire à l'imitation vocale adaptative, et ne comporte aucun élément biologique spécifique de la parole humaine. Etant donné que le phénomène de cristallisation du système se réalise pour une large partie de l'espace des paramètres de ce modèle, cela montre que l'innovation biologique permettant de passer de systèmes de vocalisations inarticulés à des systèmes caractérisés par plusieurs des propriétés fondamentales de la parole de l'humain contemporain a pu être modeste. Il ne semble en effet pas nécessaire que des structures neuronales encodant a priori et spécifiquement l'organisation phonémique, ainsi que les régularités typiques de la parole, soient générées de manière innées pour permettre la formation et l'apprentissage de tels codes de la parole. C'est la seconde contribution de ce travail : il permet de comprendre comment les propriétés d'auto-organisation de structures neuronales simples ont pu contraindre l'espace des formes biologiques de la parole et permettre qu'elles soient générées puis sélectionnées lors de la phylogenèse.

Ces nouvelles hypothèses n'auraient probablement pas pu être identifiées sans l'utilisation de simulations informatiques, parce que les dynamiques qu'elles impliquent sont complexes et difficiles à prévoir par un travail uniquement verbal. Cela illustre l'importance potentielle que peuvent avoir ces nouveaux outils méthodologiques pour les sciences humaines et pour les sciences naturelles. Cependant, ces modèles informatiques font abstraction de nombreux mécanismes biologiques et comportementaux, et donc consistent avant tout en un travail théorique de réflexion sur l'espace des

¹ Brenowitz, E.A., Beecher, M.D. (2005) Song learning in birds : diversity and plasticity, opportunities and challenges, *Trends in Neuroscience*, 28(3), pp. 127—132.

² Tyack, P. (1981) Interactions between singing hawaian humpback whales and conspecifics nearby, *Behavioral ecology and sociobiology*, 8(2), pp. 105—116.

³ Ce terme a été introduit par Gould, S. J., and Vrba, E.S. (1982) Exaptation — a missing term in the science of form, *Paleobiology* 8 (1): 4—15.

hypothèses : une fois cet espace reconceptualisé, et la cohérence interne des hypothèses évaluée par des simulations, tout le travail de validation et d'ancrage de ces hypothèses dans la réalité des observations biologiques reste à faire. Ainsi, la troisième contribution de ce travail est, plus que l'élaboration de nouvelles hypothèses spécifiques, la mise en place d'un cadre de travail et d'outils qui permettent de développer de nouvelles intuitions et de nouveaux concepts pour penser les origines et l'évolution de la parole.

Remerciements: Ce travail a été en grande partie réalisé au Sony Computer Science Laboratory, à Paris, et bénéficié du support de Luc Steels.