Cahiers Romans de Sciences Cognitives, Cuadernos Romances de Ciencias Cognitivas, Cadernos Românicos em Ciências Cognitivas, Quaderni Romanzi di Scienze Cognitive
In Cognito (2005), Vol 2 (n° 2), 1-1

AUX SOURCES DU LANGAGE : L'AUTO-ORGANISATION DE LA PAROLE

Pierre-Yves Oudeyer

Sony CSL Paris
e-mail: py@csl.sony.fr
http://www.csl.sony.fr/~py

Résumé

La parole, système conventionnel de vocalisations, est un support physique du langage qui nous permet de véhiculer des informations. Un tel système est un pré-requis pour la communication linguistique. Comment les premiers codes de la parole ont-ils donc pu apparaître sans qu'il n'existe déjà de système linguistique ? En outre, le code de la parole humain est caractérisé par des propriétés particulières : il est discret et combinatorial, il est partagé par tous les membres d'une même communauté linguistique mais peut être très différent d'une communauté à l'autre, et il existe des régularités statistiques à l'échelle de l'ensemble des langues humaines. Quelle est l'origine de cette structure ?
Pour attaquer ces questions, nous présentons dans cet article un système artificiel qui permet de conceptualiser la manière dont une société d'agents, dotés de conduits vocaux et d'oreilles reliés par des réseaux neuronaux, peut former par auto-organisation un code de la parole discret, combinatorial et partagé par tous les agents, sans que l'on présuppose de capacité linguistique ou de capacité de coordination sociale. Les structures neurales que l'on utilise sont aussi très simples d'un point de vue évolutionnaire, et leur auto-organisation repose sur le couplage de la perception et de la production à la fois dans les agents et entre les agents. Le système artificiel nous permet de mieux comprendre comment la parole a pu apparaître, en montrant comment des mécanismes d'auto-organisation ont pu contraindre l'espace des formes et aider la sélection naturelle à trouver les fondations de la parole.

Mots clés : origine de la parole, auto-organisation, évolution, formes, systèmes artificiels, agents, phonétique, phonologie

Abstract

AT THE ORIGINS OF LANGUAGE: THE SELF-ORGANIZATION OF SPEECH

Speech, as a conventional vocalization system, is a physical support for language that allows us to transmit information. Such a system is a prerequisite for linguistic communication. How then could the first speech codes have appeared without a linguistic system already in place? Furthermore, the human speech code is characterized by particular properties: it is discrete and combinatorial; it is shared by all members of the same linguistic community, but can be very different from one community to another; and there are statistical regularities, considering human languages as a whole. What is the origin of this structure?
To address these questions, in this paper we present an artificial system that allows us to conceptualize the way in which a society of agents, equipped with vocal tracts and ears connected by neural networks, can develop a shared discrete combinatorial speech code through self-organization, without presupposing linguistic capacities or the capacity for

social coordination. The neural structures that we use are evolutionarily very simple, and their self-organization relies on the coupling of perception and production both within and across agents. The artificial system allows us to better understand how speech could have appeared, by showing how self-organization mechanisms could have constrained the set of forms and helped natural selection to find the foundations of speech.

Keywords: origin of speech, self-organization, evolution, forms, artificial systems, agents, phonetics, phonology

Resumo

AS ORIGENS DA LINGUAGEM : A AUTO-ORGANIZACAO DA FALA

A fala, enquanto sistema convencional de vocalizações, é um suporte físico da linguagem que nos permite veicular informação. Tal sistema é um pré-requisito para a comunicação lingüística. De que forma então o primeiro código da fala pode ter aparecido sem que já existisse um sistema comunicação lingüístico? Além disso, o código da fala humana é caracterizado por propriedades particulares: é discreto e combinatório; é partilhado por todos os membros da mesma comunidade, embora diferente em comunidades distintas ; e há regularidades estatísticas ao se considerar o conjunto das línguas humanas. Quais as origens dessa estrutura?
Para lidar com essas questões, apresentamos neste artigo um sistema artificial que nos permite conceitualizar a maneira pela qual agentes, equipados com um trato vocal e uma orelha e conectados por redes neurais, podem auto-organizar um código de fala discreto e combinatório compartilhado. Isso é feito sem pressupor capacidades lingüísticas ou de coordenação social. As estruturas neurais usadas são simples em termos evolutivos e sua auto-organização se fundamenta no acoplamento entre produção e percepção dentro de e entre agentes. O sistema artificial permite entender melhor como a fala poderia ter aparecido, ao mostrar como a auto-organização pode restringir o espaço de formas e ajudar a seleção natural a encontrar os fundamentos da fala.

Palavras-chave: origens da fala, auto-organização, evolução, formas, sistemas artificiais, agentes, fonética, fonologia

Resumen

TITLE

El habla, como sistema convencional de vocalizaciones, es un soporte físico del lenguaje que nos permite vehicular información. Tal sistema es un pre-requisito para la comunicación lingüística. ¿De qué forma entonces el primero código del habla puede haber aparecido sin que ya existiese un sistema de comunicación lingüístico? Además, el código del habla humano es caracterizado por propiedades particulares: es discreto y combinatorio; es compartido por todos los miembros de la misma comunidad, aunque diferente en comunidades distintas; y tiene regularidades estadísticas al considerarse el conjunto de las lenguas humanas. ¿Cuáles son los orígenes de esa estructura?
Para tratar estas preguntas, presentamos en este artículo un sistema artificial que nos permite conceptualizar la manera por la cual agentes equipados con un trato vocal y un oído, y conectados por redes neuronales, pueden auto-organizar un código de habla discreto y combinatorio compartido. Aquello se hace sin presuponer capacidades lingüísticas o de coordinación social. Las estructuras neuronales empleadas son simples en términos evolutivos y su auto-organización se basa en el acoplamiento entre producción y percepción dentro de y entre agentes. El sistema artificial permite entender mejor como el habla podría

haber aparecido, enseñando como la auto-organización puede restringir el espacio de formas e ayudar la selección natural a encontrar los fundamentos del habla.

Palabras-clave: orígenes del habla, auto-organización, evolución, formas, sistemas artificiales, agentes, fonética, fonología.

Estratto

LE ORIGINI DEL LINGUAGGIO : L'AUTO-ORGANIZZAZIONE DEL LINGUAGGIO PARLATO

La parola, come sistema convenzionale di vocalizzazioni, è un supporto fisico del linguaggio che permette la trasmissione dell'informazione. Tale sistema costituisce un prerequisito per la comunicazione linguistica. In che modo i primi codici del linguaggioparlato sono potuti apparire in assenza di un sistema linguistico preesistente? Inoltre, il codice del linguaggio parlato umano è caratterizzato da alcune proprietà particolari: esso è discreto e combinatorio, è condiviso da tutti i membri di una stessa comunità linguistica, ma puó variare molto da una comunità all'altra; ed esistono delle regolarità statistiche che accomunano le lingue umane. Qual è l'origine di questa struttura?
Per affrontare tali interrogativi, presentiamo in questo articolo un sistema artificiale che permette di concettualizzare il modo in cui una società di agenti, dotati di condotti vocali e uditivi collegati tramite delle reti neurali, puó auto-organizzarsi in modo da formare un codice del linguaggio parlato discreto, combinatorio e condiviso da tutti gli agenti, senza presupporre alcuna abilità linguistica o capacità di coordinazione sociale preesistente. Le strutture neurali qui utilizzate sono alquanto semplici anche dal punto di vista evolutivo, e la loro auto-organizzazione si fonda sull'abbinamento della percezione e della produzione all'interno degli agenti e tra gli agenti stessi. Tale sistema artificiale ci permette di comprendere meglio come il lainguaggio parlato puó determinarsi, mostrando in che modo dei meccanismi di auto-organizzazione hanno potuto restringere lo spazio delle forme ed aiutare la selezione naturale a trovare i fondamenti della parola.

Parole-chiave : origini del parlato, auto-organizzazione, evoluzione, forme, sistemi artificiali, agenti, fonetica, fonologia

Les origines du langage : un champ de recherche florissant

Il y a très longtemps, les humains ne produisaient que des grognements inarticulés. Or, maintenant ils parlent. La question de savoir comment ils en sont venus à parler est l'une des questions les plus difficile qui soit posée à la science. Alors qu'elle a été éludée de la scène scientifique pendant la presque totalité du 20ème siècle, à la suite de la déclaration de la Société Linguistique de Paris qui la bannit de sa constitution, elle est redevenue le centre des recherches de toute une communauté de chercheurs. La diversité des problématiques qui sont impliquées induit une forte pluridisciplinarité : des linguistes, des anthropologues, des spécialistes de neurosciences, des primatologistes, des psychologues, mais aussi des physiciens et des informaticiens. En effet, l'un des grands axes théoriques de la recherche sur les origines du langage considère qu'un certain nombre de propriétés du langage ne s'expliquent que par la dynamique des interactions complexes des entités qui sont impliquées (les interactions entres les circuits neuronaux, le conduit vocal, l'oreille, mais aussi les interactions des individus qui les portent dans un environnement réel). C'est l'apport de la théorie de la complexité (Nicolis et Prigogine, 1977), développée au 20ème siècle, qui nous a appris qu'il y a de nombreux systèmes naturels dans

lesquels les propriétés macroscopiques sont irréductibles aux propriétés microscopiques. C'est ce qu'on appelle l'auto-organisation. C'est par exemple le cas des structures fascinantes des nids de termites, dont la forme n'est ni codée ni connue par aucune des termites prises individuellement, mais apparaît de manière auto-organisée lors de leurs interactions. C'est aussi le cas de la formation des cristaux de glace à partir de molécules d'eau, comme l'illustre la figure 1. Or ces phénomènes d'auto-organisation sont souvent compliqués à comprendre ou à prévoir intuitivement, et à formuler verbalement.

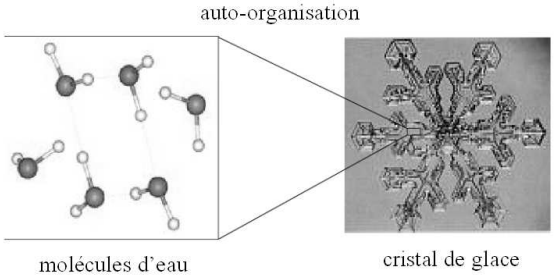

Figure 1 : Le phénomène d'auto-organisation : les propriétés globales du cristal de glace sont qualitativement différentes des propriétés locales des molécules d'eau

La méthode de l'artificiel

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 la recherche des origines du langage.

En effet, ils disposent d'outils nouveaux et indispensables pour nous permettre de comprendre les phénomènes d'auto-organisation dans les systèmes complexes. Ils construisent des modèles opérationnels des interactions entre les sous-systèmes impliqués dans le langage. Un modèle opérationnel est un système qui définit formellement l'ensemble de ses pré-suppositions 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 ainsi que 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 de modéliser des phénomènes plus complexes que le premier, est celui utilisé par les chercheurs en intelligence artificielle: 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, qu'on pourra appeler robots même s'ils évoluent dans des environnements virtuels, dotés de cerveaux et de corps artificiels. Ceux-ci 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 » (Steels, 1997).

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 composants. 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 bien souvent on découvre des failles dans les pré-suppositions ainsi que dans les conclusions qui doivent être révisées); 2) ils servent à explorer et générer de nouvelles théories, 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 obtenu et ont permis d'ouvrir la voie à la résolution de questions jusque là sans réponses : la génération décentralisée de conventions lexicales et sémantiques dans des communautés de robots (Steels, 1997; Kaplan, 2001), avec par exemple l'expérience des « Talking Heads » (http://talking-heads.csl.sony.fr), 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 humaines (de Boer, 2001; Oudeyer, 2001), avec par exemple l'expérience intitulée « Maîdo et Gurby » (http://www.csl.sony.fr/~py/videos.html), la formation de structures syntaxiques conventionnalisées (Batali, 1998) ou les conditions dans lesquelles la compositionalité peut être sélectionnée (Kirby, 1998).

Il est important de noter que cette méthodologie de l'artificiel, dans le cadre de la recherche sur les origines du langage, est avant tout une méthodologie exploratoire. Elle s'insère dans une logique scientifique d'abduction. Le mot « modèle » qui est souvent employé dans la littérature sur les origines du langage pour décrire les systèmes artificiels a un sens différent de son acceptation traditionnelle. C'est pourquoi d'ailleurs je préfère tout simplement employer l'expression « système artificiel ». En effet, il ne s'agit pas d'observer un phénomène naturel et d'essayer d'en abstraire les mécanismes et les variables fondamentales pour construire un modèle qui soit capable de prédire précisément la réalité. 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 de la construction de système artificiel : développer notre intuition sur les dynamiques de formation du langage, et ébaucher l'espace des hypothèses. Il ne s'agit donc pas d'établir directement 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.

Pour ne pas que cela reste abstrait, nous allons maintenant détailler les grandes lignes d'un exemple de système artificiel construit dans le but de faire progresser la réflexion et la conceptualisation des origines du langage. Cet exemple ne s'attaque pas au problème de l'origine du langage dans sa généralité, mais plutôt à la question de l'origine de l'un de ses composants essentiels : la parole, c'est à dire les systèmes

de sons, véhicules et supports physiques du langage (au même titre par exemple les signes visuels dans les langues des signes).

Le code de la parole

Les humains ont un système de vocalisations complexe. Celles-ci sont digitales et compositionnelles : elles sont construites à partir de la re-combinaison d'unités qui sont systématiquement ré-utilisées dans les vocalisations. Ces unités sont présentes à plusieurs niveaux (e.g. les gestes, les coordinations de gestes ou phonèmes, les morphèmes). Alors que l'espace articulatoire qui définit l'espace des gestes est continu (voir figure 2), chaque langue discrétise cet espace à sa manière. Alors qu'il y a une grande diversité dans l'ensemble des répertoires de ces unités dans les langues du monde, il y a en même temps de fortes régularités (par exemple, la fréquence élevée du système à cinq voyelles /e, i, o, a, u/). La manière dont les unités sont combinées est aussi très particulière : 1) toutes les séquences de phonèmes ne sont pas autorisées dans une langue donnée, 2) l'ensemble des combinaisons de phonèmes est organisé en patterns.

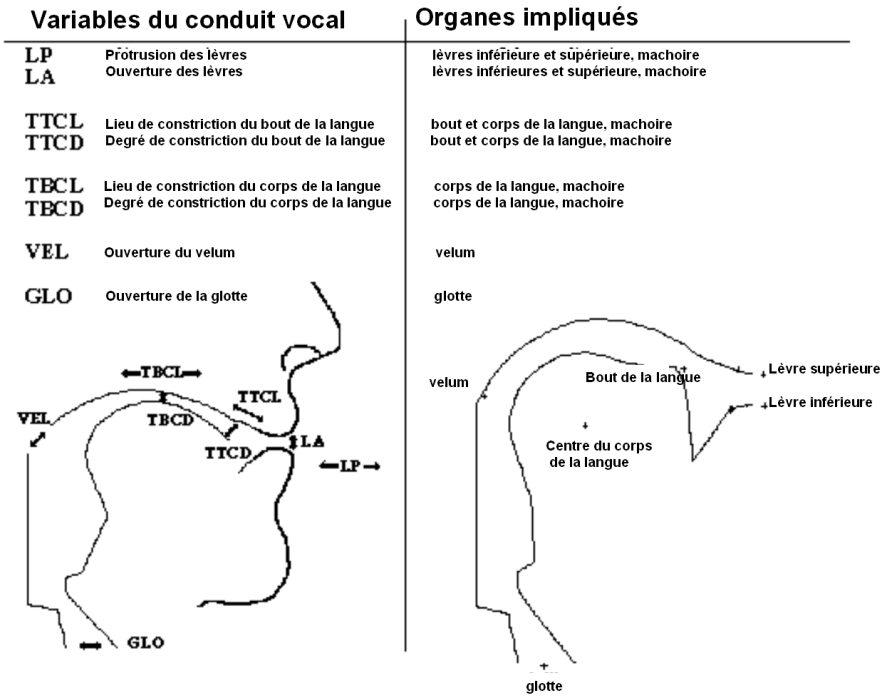

Figure 2 : Les vocalisations sont produites par des mouvements des organes articulateurs, et sont donc physiquement définies dans un espace moteur continu. Cependant, chaque langue discrétise cet espace à sa manière.

Cette organisation en pattern veut dire que par exemple, on peut résumer les combinaisons de phonèmes autorisées en Japonais pour former des syllabes (« moras » plus exactement) par les patterns « CV/CVC/VC », où par exemple

« CV » est un pattern 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.

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 ? Il est vrai que depuis par exemple les travaux de de Boer (2001) ou Kaplan (2001), 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 « de dynamique du consensus », font appel à la pré-existence de conventions et d'interactions linguistiques. Ils concernent donc plutôt l'évolution des langues, mais ne proposent pas de solution quant 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 ?

Comment sont apparus les premiers codes de la parole, ou comment l'auto-organisation peut aider la sélection naturelle ?

Il est ainsi naturel de se demander d'où vient cette organisation et comment un tel code conventionnel et partagé a pu se former dans une société d'agents qui ne disposaient pas déjà de conventions. 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 Lindblöm (1992) qui a montré que les régularités statistiques des répertoires de phonèmes pouvaient être prédites en recherchant quels étaient 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 se révèle pas suffisamment efficace pour trouver des structures complexes comme celles de la parole : l'espace de recherche est trop grand (Ball, 2001). C'est pourquoi un second type de réponse est nécessaire : il faut aussi établir comment la sélection naturelle a pu trouver ces structures. On peut pour cela montrer comment l'auto-organisation a pu contraindre l'espace de recherche et aidé la sélection naturelle. Cela peut être fait en montrant qu'un système beaucoup plus simple s'auto-organise spontanément en formant la structure que l'on cherche à expliquer. En fait, nous reprenons pour la question de l'origine de la parole la même structure argumentative que celle de D'Arcy Thompson

(1932) à propos de l'explication des formes hexagonales des cellules de cires dans les ruches des abeilles[1] (voir figure 3).

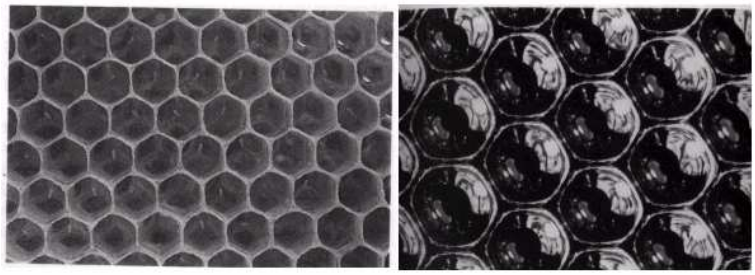

Figure 3 : L'auto-organisation des formes hexagonales des cellules de cires dans les ruches des abeilles

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.

Le système artificiel

Nous allons présenter dans ce paragraphe un résumé de l'architecture du système, et en particulier de l'architecture des agents. Les détails techniques sont présentés dans l'annexe 1.

---

[1] Les cellules dans les ruches des abeilles ont une forme hexagonale parfaite. Comment les abeilles en sont-elles venues à bâtir de telles structures ? Un premier élément de réponse apparaît si l'on remarque que l'hexagone est la forme qui nécessite le moins de cire pour couvrir un plan avec des cellules d'une surface donnée. Donc l'hexagone permet aux abeilles d'économiser de l'énergie métabolique, ce qui leur permet de mieux survivre et de se répliquer plus efficacement que si elles construisaient d'autres formes. On peut donc fournir l'explication néo-darwinienne classique suivante : les abeilles ont du commencer par construire des formes aléatoires, puis par le jeu des mutations aléatoires et de la sélection naturelle, petit à petit, les formes plus efficaces ont été sélectionnées, jusqu'à ce qu'on en arrive à l'hexagone parfait, forme optimale. Maintenant, il se trouve qu'un génome qui conduirait à des abeilles qui construisent exactement des hexagones, doit être assez complexe et est pour le moins une aiguille dans une botte de foin ! Or, il semble que la version classique du mécanisme néo-darwinien avec mutation aléatoire/sélection ne soit pas de manière évidente assez efficace pour avoir permis à la sélection naturelle d'avoir trouvé un tel génome. L'explication n'est donc pas suffisante. D'Arcy Thompson l'a complétée. Il s'est aperçu que lorsque des cellules de cire de forme pas trop tordues étaient chauffées comme elles le sont par le battement des ailes des abeilles, elles ont à peu près les mêmes propriétés physiques que des gouttes d'eau qu'on entasse les unes sur les autres. Et justement, il se trouve que quand on entasse des gouttes d'eau les unes sur les autres, elles prennent spontanément la forme d'hexagones ! Ainsi, D'Arcy Thompson montre que la sélection naturelle n'a pas eu besoin de trouver des génomes qui pré-programmaient précisément la construction d'hexagones, mais seulement des génomes qui faisaient construire aux abeilles des cellules à peu près rondes, pas trop tordues, et à peu près de la même taille, et que la physique faisait le reste ! Il a ainsi montré comment des mécanismes physiques auto-organisés (bien que le terme n'existait pas encore) pouvaient contraindre l'espace des formes et faciliter grandement le travail de la sélection naturelle.

Le système artificiel est basé sur le couplage de dispositifs nerveux sensori-moteurs génériques qui sont câblés aléatoirement au départ et implémentés dans la tête des agents artificiels. Les agents disposent d'une oreille artificielle, capable de transformer un signal acoustique en impulsions nerveuses qui stimulent les neurones d'une carte de neurones artificiels perceptuels. Les agents disposent aussi d'une carte de neurones moteurs, qui peuvent être activés soit par les neurones perceptuels soit spontanément, et dont l'activation, couplée à la génération d'un signal interne de déclenchement du mouvement, envoie des signaux à un contrôleur qui fait bouger les organes d'un conduit vocal artificiel. Ces signaux des neurones moteurs correspondent à des commandes qui spécifient des objectifs articulatoires à atteindre dans un timing donné. Les objectifs articulatoires sont typiquement définis comme des relations entre certains organes du conduit vocal (comme la distance entre les lèvres ou le lieu de constriction de la langue). Ils sont une implémentation du concept de « gestes » développé par la phonologie articulatoire (Browman et Goldstein, 1986). La figure 4 présente une vue générale de cette architecture. Dans cet article, nous considérerons des espaces de relations entre organes à deux ou trois dimensions (e.g., lieu et manière de constriction de la langue et rondeur des lèvres pour la production des voyelles).

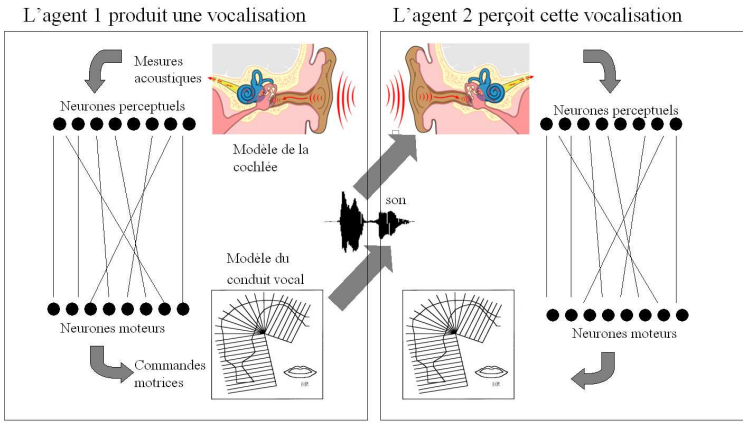

Figure 4 : Architecture du système artificiel

Ce que nous appelons ici « neurone » est une boite qui reçoit en entrée un certain nombre de mesures/signaux, les intègre et calcule son niveau d'activation qui est à son tour propagé vers l'entrée d'autres neurones grâce à des connexions de sortie. Typiquement, l'intégration est réalisée en pondérant chacune des entrées (i.e., en multipliant ces entrées par un poids), puis en les sommant et en appliquant au résultat une fonction de transfert, que l'on appelle aussi fonction d'activation. La fonction d'activation est ici une gaussienne, dont la largeur est un paramètre du système. Un poids est attaché à chaque connexion entre deux neurones. Dans le système, tous les poids sont initialement petits et aléatoires.

En outre, tous les neurones dans chaque carte neurale (perceptuelle et motrice) sont totalement interconnectés. Cela veut dire qu'ils reçoivent des entrées de tous les neurones dans la même carte nerveuse. Quand un stimulus sonore est perçu, cela donne lieu à une première activation de tous les neurones des deux cartes neurales. Cette activation initiale est la base d'une mise à jour de tous les poids des

connexions que l'on décrira plus loin. Une fois cette mise à jour effectuée, l'activation de chaque neurone est elle-même mise à jour en fonction de l'activation des autres neurones de sa carte et grâce aux connexions qui les relient. Cette mise à jour se répète jusqu'à ce que l'ensemble des activations se stabilise, ce qui s'appelle un point-attracteur dans le langage des sciences des systèmes dynamiques. Cet attracteur est le même pour tout un ensemble de stimuli initiaux, que l'on appelle son bassin d'attraction. Cela permet de modéliser un comportement de catégorisation. Il y a autant de catégories sonores qu'il y a d'attracteurs.

La production d'une vocalisation consiste en l'activation aléatoire d'une séquence de plusieurs neurones moteurs. L'activation de chaque neurone moteur envoie au contrôleur du conduit vocal l'instruction d'atteindre la configuration articulatoire codée par les poids de ses connections de sortie. Le contrôleur génère alors une trajectoire continue, correspondant à un mouvement des organes articulatoires. Cela est réalisé ici simplement en calculant une interpolation polynomiale. Au départ, les poids de sortie des neurones moteurs sont aléatoires et répartis uniformément dans l'espace des configurations articulatoires possibles. Cela veut dire que les vocalisations qu'ils produisent au départ sont holistiques car les objectifs articulatoires qui sont utilisés sont répartis uniformément sur tous l'espace continu des configurations possibles, et il n'y a pas de ré-utilisation organisée d'objectifs articulatoires d'une vocalisation à l'autre. En résumé, leurs vocalisations ne sont pas caractérisées par le codage phonémique au départ.

Les agents produisent des vocalisations qui correspondent à un mouvement continu du conduit vocal, et non pas à des configurations articulatoires statiques. Cela implique que les agents perçoivent les vocalisations des autres comme des trajectoires continues dans l'espace acoustique. Nous allons expliquer maintenant comment ces trajectoires sont traitées et comment leur perception modifie les poids des connections des neurones artificiels.

Tout d'abord, les agents ne sont pas capables de détecter des évènements de « haut niveau » dans une trajectoire acoustique, et en particulier ne sont pas capables d'inférer quels sont les points de cette trajectoire qui correspondent aux objectifs articulatoires qui ont été utilisés par l'agent qui l'a produite. Au contraire, les agents segmentent la trajectoire en petits morceaux très courts, correspondant à la résolution temporelle de la cochlée. Ensuite, chacune de ces petites parties est moyennée, ce qui donne un point dans l'espace acoustique qui est envoyé comme stimulus en entrée de la carte neurale perceptuelle, ce qui active les neurones perceptuels. L'activation est alors propagée aux neurones moteurs, et les deux cartes sont mises à jour comme on va l'expliquer. Puis, le point moyenné suivant de la trajectoire est envoyé à la suite à la carte perceptuelle, les mises à jours sont faites, et on recommence ce processus jusqu'à ce qu'on arrive au bout de la trajectoire acoustique.

La mise à jour des poids des connections des neurones perceptuels a lieu chaque fois que les neurones sont activés. Les connections d'entrée des neurones perceptuels sont modifiées de manière à ce que les neurones deviennent plus sensibles au stimulus qui les a activés, et la modification est d'autant plus grande que l'activation du neurone est grande.

Pour la mise à jour des neurones moteurs, deux cas se présentent : 1) les neurones moteurs sont déjà activés au moment où la carte perceptuelle est activée parce que la vocalisation a été produite par l'agent lui-même, et alors les poids des connexions entre les neurones perceptuels et moteurs sont renforcés si elles relient des

neurones dont les activités sont corrélées, et affaiblies si elles relient des neurones dont les activités ne sont pas corrélées (ç'est une loi d'apprentissage hebbienne). Cette loi d'apprentissage permet aux agents d'apprendre les correspondances entre des stimuli acoustiques et les commandes motrices qui les ont produites pendant le babillage. 2) Si les neurones moteurs ne sont pas déjà activés quand les neurones perceptuels le sont, ç'est à dire quand la vocalisation a été produite par un autre agent, alors les poids des connections qui relient les neurones perceptuels et les neurones moteurs ne sont pas modifiés. Cependant, l'activation des neurones perceptuels est propagée aux neurones moteurs par le biais de ces connexions, et les poids des connexions entre les neurones moteurs et le contrôleur du conduit vocal sont modifiés. Le neurone moteur qui a la plus haute activation est sélectionné, et les poids de ses connexions de sortie, qui spécifient une relation entre organes, sont utilisés comme référence pour mettre à jour les autres poids : ils sont modifiés de tel manière que la relation entre organes qu'ils spécifient ressemblent un peu plus à la relation « référence », et cette modification est pondérée par l'activation courante de chaque neurone.

Une caractéristique cruciale de l'architecture est le couplage entre les processus de production et les processus de perception. Dans ce qui suit, nous appellerons « vecteur préféré » les poids des connexions d'entrées des neurones perceptuels. Ce nom vient du fait que l'ensemble de ces poids forme un vecteur, et que le stimulus qui a les mêmes valeurs est celui qui active maximalement le neurone. Nous appellerons aussi « vecteur préféré » les poids de sortie d'un neurone moteur. L'architecture et la dynamique du système artificiel est telle que la distribution des vecteurs préférés des neurones moteurs se cale sur celle des vecteurs préférés des neurones perceptuels : si l'on active au hasard les neurones de la carte motrice et que les sons correspondants sont prononcés, alors cela donne une distribution de son qui va tendre à être la même que celle qui est encodée par les neurones de la carte perceptuelle. La distribution des vecteurs préférés des neurones perceptuels est modifiée quand des sons sont perçus : cela implique que si un agent entend certains sons plus souvent que d'autres, il aura tendance à produire ces sons plus souvent que les autres (ici le terme « son » désigne une partie moyennée d'une trajectoire acoustique générée par le filtre de résolution temporelle que l'on a décrit plus haut). Il est intéressant de noter que ce processus de « calage des distributions » n'est pas réalisé au travers d'une imitation, mais ç'est un effet de bord de l'augmentation de la sensibilité des neurones aux stimuli et aux transferts des activations entre les cartes perceptuelles, ce qui est un mécanisme neural générique de très bas niveau (Kandel *et al.,* 2001).

Les agents sont disposés dans un environnement virtuel dans lequel ils se déplacent de manière aléatoire. A des moments choisis aléatoirement, ils produisent une vocalisation, et l'agent le plus proche entend cette vocalisation et adapte ses cartes neurales en fonction de sa perception. Chaque agent écoute aussi ses propres sons, ce qui lui permet d'apprendre les correspondances entre l'espace perceptuel et l'espace moteur.

Tous les neurones des cartes motrices et perceptuelles sont initialement aléatoires et uniformes. Cela veut dire que leurs vocalisations sont holistiques et inarticulées : l'espace continu des configurations aléatoires est utilisé uniformément. Les vocalisations sont donc holistiques. Nous allons montrer que ces cartes neurales s'auto-organisent et se synchronisent de telle manière qu'après un certain temps les agents produisent des vocalisations dont les objectifs articulatoires appartiennent à un petit nombre de clusters/modes biens définis : le continuum sera alors discrétisé.

En outre, le nombre de clusters est petit comparé au nombre de vocalisations que les agents produisent durant leur vie, ce qui implique une ré-utilisation systématique des objectifs articulatoires entre les vocalisations. Enfin, ces clusters sont les mêmes pour tous les agents : le code est partagé et spécifique à chaque communauté. En effet, à chaque fois qu'on fait tourner la simulation, on obtient un code de la parole différent : c'est la diversité linguistique.

Nous utilisons deux sortes de modèles du conduit vocal qui permettent de transformer des trajectoires articulatoires en trajectoires acoustiques. Le premier type est abstrait et linéaire : on utilise un espace moteur abstrait de dimension deux, et on transforme chacun de ses point en un point dans un espace acoustique abstrait de dimension deux grâce à une fonction linéaire. L'utilisation de ce modèle permet de déterminer quels sont les résultats qui sont dus intrinsèquement aux propriétés dynamiques du couplage des cartes sensori-motrices, et ne nécessitent pas la présence de non-linéarités dans la fonction qui transforme des articulations en ondes sonores. Nous montrerons en fait que l'on peut aller assez loin sans présupposer de non-linéarités : discrétisation, ré-utilisation systématique, partage, diversité. Nous utilisons ensuite un modèle du conduit vocal plus réaliste, qui concerne la production des voyelles : la hauteur du corps de la langue, la position de constriction du corps de la langue, et la rondeur des lèvres. Les formants correspondants à chaque configuration sont alors calculés en utilisant le synthétiseur articulatoire de de Boer (de Boer, 2001), qui a lui même été construit à partir de données sur la production des voyelles humaines. Nous allons montrer que ce modèle nous permet de prédire les systèmes de voyelles qui apparaissent le plus fréquemment dans les langues humaines.

La dynamique du système artificiel : résultats quand on utilise le modèle abstrait du conduit vocal

Les résultats que nous allons décrire concernent des simulations dans lesquels la population était de 20 agents, et dans lesquelles chaque carte de neurone est composée de 500 neurones. L'espace perceptuel et l'espace moteur sont tous les deux de dimension 2, à valeur continues entre 0 et 1.

Au départ, comme les vecteurs préférés des neurones sont répartis uniformément et aléatoirement dans l'espace, les objectifs articulatoires qui composent les vocalisations des agents sont aussi distribués uniformément et aléatoirement. La figure 5 montre les vecteurs préférés des neurones perceptuels de deux agents. On peut observer qu'ils couvrent uniformément l'espace. Il n'y a pas d'organisation. La figure 6 montre le processus de relaxation dynamique associé à ces cartes perceptuelles et dû à leurs connections récurrentes. C'est une représentation de leur comportement de catégorisation. En effet, chaque petite flèche représente le changement global d'activation après une itération de la relaxation (voir l'annexe). Le début d'une flèche représente un pattern d'activation au temps t (généré en présentant un stimulus dont les coordonnées correspondent aux coordonnées du début de la flèche). La fin de la flèche représente le *pattern* des activations de la carte neurale après une itération de la relaxation. L'ensemble de toutes les flèches permet de visualiser plusieurs itérations : on part de la flèche dont le début correspond au stimulus, et on suit les flèches. Au bout d'un certain temps, quel que soit le point de départ, on arrive à un point fixe. Cela correspond à un attracteur de la dynamique de la carte neurale, dont le bassin d'attraction définit la catégorie du

stimulus initial, et le point fixe le prototype de cette catégorie. Avec des vecteurs préférés répartis uniformément et aléatoirement comme ç'est le cas initialement, le nombre des attracteurs ainsi que les frontières de leurs bassins d'attraction initiaux sont aléatoires.

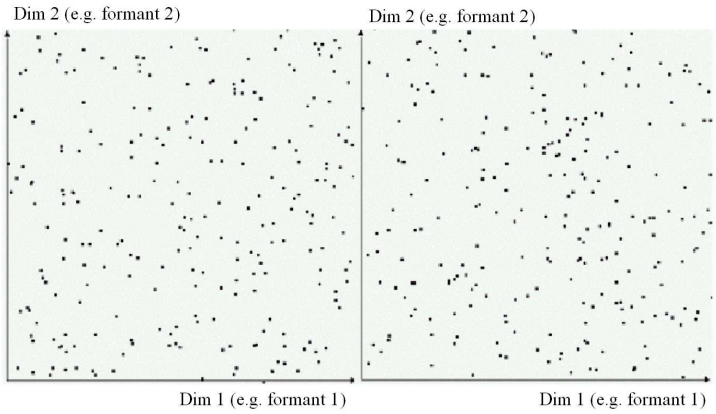

Figure 5 : Représentation des vecteurs préférés des neurones des cartes perceptuelles de deux agents au début de la simulation : on observe qu'ils sont réparties aléatoirement et uniformément.

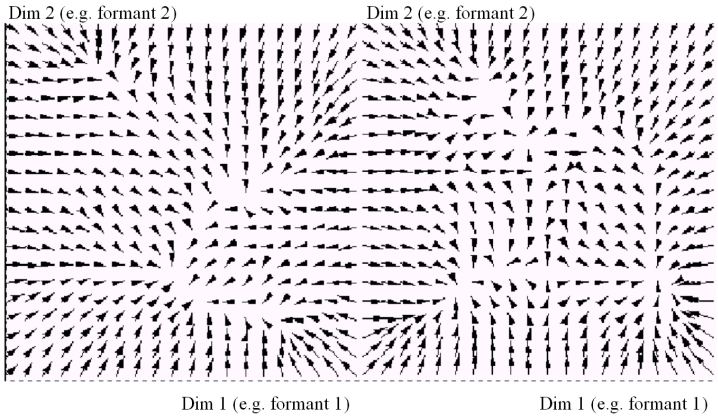

Figure 6 : Le paysage des attracteurs associés aux deux cartes neurales de la figure précédente.

Comme on l'a expliqué plus haut, la loi d'apprentissage de la carte perceptuelle est telle qu'elle tend à approximer la distribution des sons qui sont entendus par l'agent dans son environnement. Tous les agents produisent initialement des vocalisations complexes composées d'objectifs articulatoires uniformément distribués. En conséquence, cette situation initiale est un équilibre. Cependant, c'est un équilibre instable, et les fluctuations dues à la stochasticité inhérente au système assurent qu'à un moment donné la symétrie va se casser : de temps en temps, certains sons sont prononcés plus souvent que d'autres, et ces fluctuations aléatoires peuvent être amplifiées par une boucle de rétro-action positive. Cela conduit à une distribution

multi-modale : les agents se retrouvent avec des cartes neurales comme celles représentées sur la figure 7, qui sont les mêmes que celles de la figure 5 mais après 2000 interactions dans une population de 20 agents. La figure 7 montre que la distribution des vecteurs préférés n'est plus uniforme mais au contraire des clusters sont apparus. Cependant, il n'est pas facile de visualiser ces clusters sur la représentation de la figure 7, parce que tous les points de chaque *clusters* apparaissent comme un seul point et certains autres points correspondent à des neurones dont le vecteur préféré n'appartient à aucun cluster. Ces points « errants » ne sont pas représentatifs de la distribution, mais introduisent un bruit dans la représentation. La figure 8 nous permet de mieux visualiser la distribution en montrant le paysage des attracteurs qui est associé à chaque carte neurale. Nous observons qu'il y a maintenant trois attracteurs bien définis chez chaque agent, correspondant à trois clusters, et que ces clusters sont les mêmes chez les deux agents représentés (et les mêmes chez les 18 autres agents que l'on a pas représentés). Cela veut dire que les objectifs articulatoires que les agents utilisent appartiennent maintenant systématiquement à l'un ou l'autre de ces clusters, et peuvent être catégorisés automatiquement comme tels grâce au mécanisme de relaxation du réseau. Le continuum articulatoire est maintenant discrétisé. En outre, le nombre de clusters qui est apparu est petit, ce qui conduit à la ré-utilisation systématique des objectifs articulatoires dans la construction des vocalisations. Tous les agents partagent le même code de la parole à la fin de chaque simulation, mais ce code est différent à chaque nouvelle simulation. Avec exactement les mêmes paramètres de la simulation, on obtient des nombres de modes/catégories et des prototypes qui leurs sont associés qui varient. Cela est dû à la stochasticité intrinsèque au système. Nous illustrerons ce phénomène dans la partie suivante.

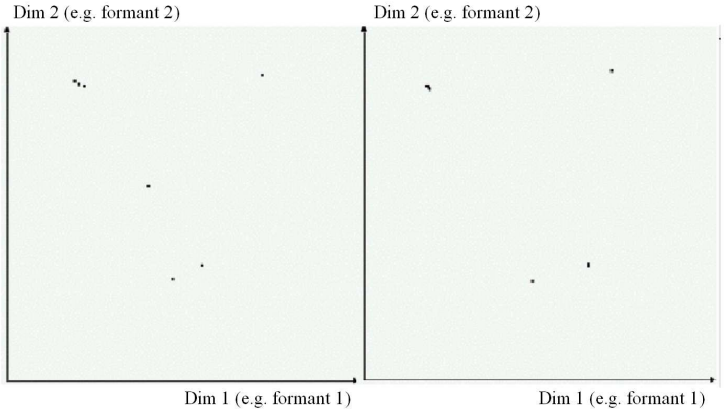

Figure 7 : Représentation des distributions des vecteurs préférés des deux cartes de la figure 5 après 2000 interactions vocales : on observe que le continuum a été discrétisé.

Il faut noter que ce phénomène de cristallisation est valable pour n'importe quel nombre d'agents (expérimentalement), et en particulier avec un seul agent qui s'adapte à ses propres vocalisations. Cela veut dire que l'interaction avec les autres agents – i.e., la composante sociale – n'est pas nécessaire pour l'apparition de la discrétisation et de la ré-utilisation systématique des objectifs articulatoires. Ce qui est intéressant, c'est que quand les agents interagissent, ils se cristallisent et se synchronisent sur le même système de catégories sonores. Il y a donc deux résultats

indépendants : d'une part, la discrétisation et la ré-utilisation automatique apparaissent grâce au couplage entre la perception et la production dans les agents, et d'autre part le partage des catégories phonémiques apparaît grâce au couplage entre la perception et la production entre les agents.

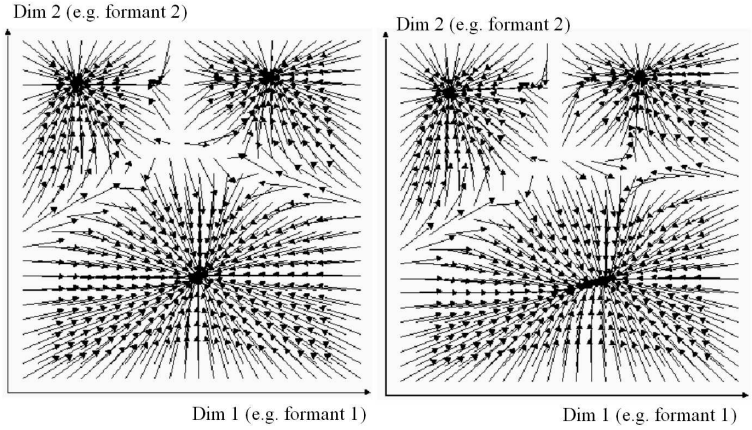

Figure 8 : Représentation du paysage des attracteurs associés aux deux cartes neurales perceptuelles de la figure précédente. On observe la présence de trois bassins d'attractions, correspondants à trois clusters majeurs et donc à trois catégories phonémiques.

Nous observons aussi que les attracteurs qui apparaissent sont relativement bien répartis au travers de l'espace. Les prototypes que leurs centres définissent sont donc perceptuellement assez distincts. Dans les termes définis par le travail de Lindblöm (Lindblöm, 1992), l'énergie de ces systèmes est élevée. Cependant, il n'y a pas de pression fonctionnelle pour éviter que les prototypes sonores soient trop proches. Ils sont distribués de cette manière grâce aux propriétés intrinsèques des réseaux de neurones récurrents et leurs fonctions d'activation gaussienne assez large : en effet, si deux clusters de neurones sont trop près, alors la sommation de leurs fonction d'activation dans le processus de relaxation lisse localement la distribution des neurones et un seul attracteur apparaît.

La dynamique du système artificiel : résultats quand on utilise le modèle réaliste de la production des voyelles

Dans la partie précédente, nous avons utilisé un système de conduit vocal abstrait qui transformait des configurations articulatoires en signaux acoustiques par une simple fonction linéaire. En d'autres termes, nous ne prenions pas en compte les contraintes du conduit vocal humain dues à sa non-linéarité. C'était utile car nous avons pu montrer qu'aucune asymétrie initiale dans le système n'était nécessaire pour qu'apparaissent une discrétisation du continuum articulatoire. Cela montre qu'en principe il n'y a pas besoin de discontinuités ou de non-linéarités dans la fonction qui fait correspondre des signaux acoustiques à des configurations articulatoires afin d'expliquer l'existence de la digitalité de la parole. Cependant, cela n'implique pas que des non-linéarités ne peuvent pas y contribuer.

Cependant, une fonction de transformation vocale réaliste a des propriétés particulières qui introduisent des biais dans la formation des structures de la parole. En effet, avec le conduit vocal humain, il y a des configurations articulatoires pour lesquelles une petite modification conduit à une petite modification du son produit, et d'autres configurations articulatoires pour lesquelles une petite configuration articulatoire conduit à de grandes modifications acoustiques. Alors que les neurones dans la carte neurale motrice ont au départ des vecteurs préférés qui ont une distribution uniforme, cette distribution va vite être biaisée : à cause des non-linéarités et des règles d'apprentissage, la production de certains sons va avoir plus d'influence que la production d'autres sons. Pour certains d'entre eux, beaucoup de neurones moteurs verront leur vecteur préféré modifié substantiellement, alors que pour d'autres sons, seuls quelques neurones moteurs seront affectés. Cela conduit très rapidement à des non-uniformités dans la distribution des vecteurs préférés dans les cartes motrices, avec plus de neurones dans les parties de l'espace moteur pour lesquelles de petites modifications conduisent à des petites modifications du son correspondant, et moins de neurones dans les parties de l'espace moteur pour lesquelles de petites modifications conduisent à de grandes modifications acoustiques. En conséquence, la distribution des objectifs articulatoires qui composent les vocalisations sera biaisée, et le mécanisme d'adaptation des neurones perceptuels conduira aussi à un biais de la distribution de leurs vecteurs préférés.

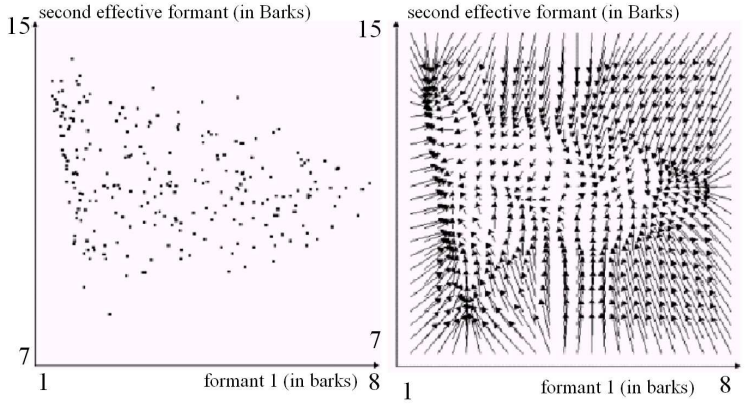

Figure 9 : Représentation de la carte perceptuelle d'un agent qui utilise le modèle réaliste de production des voyelles, et après 200 interactions. Sur la gauche : la distribution des vecteurs préférés. Sur la droite : le paysage des attracteurs associés. On observe que la distribution des vecteurs préférés est biaisée, en conséquence des non-linéarités de la fonction qui fait correspondre des sons à des configurations articulatoires.

Nous utilisons donc dans cette partie un synthétiseur articulatoire qui a été développé par de Boer (de Boer, 2001). Celui-ci modélise la production des voyelles. Le fait que les agents produisent uniquement des vocalisations composées de voyelles n'implique pas que le système ne s'applique pas aux sons consonnes. Nous avons choisi ce synthétiseur articulatoire parce que c'était le seul qui est à la fois réaliste et rapide pour effectuer plusieurs milliers de simulations. L'espace articulatoire a ici trois dimensions : la hauteur du corps de la langue (la manière

d'articulation), la position du corps la langue (le lieu d'articulation), et la rondeur des lèvres. Chaque jeu de valeur pour ces paramètres est transformé en une représentation acoustique correspondant aux quatre premiers formants. Ensuite, le second formant effectif est calculé comme une combinaison non-linéaire des quatre formants. Le premier formant et le second formant effectif sont connus pour être de bons modèles de notre perception des voyelles (de Boer, 2001). Pour en avoir une idée, la figure 9 montre l'état des cartes neurales acoustiques d'un agent après 200 interactions (soit peu après le départ de la simulation). On peut observer le biais dans la distribution des vecteurs préférés des neurones dus aux non-linéarités.

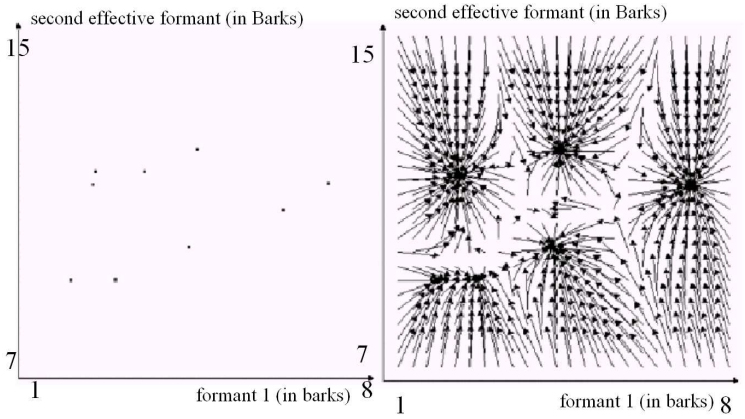

Figure 10 : Un exemple de carte neurale perceptuelle obtenue après cristallisation du système et correspondant au système de voyelle le plus fréquent à la fois dans les simulations et dans les langues humaines.

Une série de 500 simulations a été effectuée avec les mêmes jeux de paramètres, et à chaque fois le nombre de voyelles ainsi que la structure des systèmes de voyelles émergents ont été mesurés. Chaque système de voyelle a été classé en fonction de la position relative des voyelles, et non pas en regardant la position précise de chacune d'entre elles. Cette méthode est inspirée du travail de Crothers (Crothers, 1978) sur les universaux dans les systèmes de voyelles, et est identique au type de classification effectué par de Boer (de Boer, 2001). Un premier résultat montre que la distribution de la taille des systèmes émergents est très similaire à celle des systèmes de voyelles des langues humaines : la figure 11 montre les deux distributions. On observe en particulier que dans les deux cas il y a un pic à 5 voyelles, ce qui est remarquable car 5 n'est ni le minimum ni le maximum. Cette prédiction est même un peu plus exacte que celle présentée par de Boer (de Boer, 2001) car son modèle prédisait un pic à 4 voyelles. Ensuite, on a comparé la structures des systèmes de voyelles émergents avec celle des langues humaines comme elle est décrite dans (Schwartz *et al.,* 1997). Plus précisément, la distribution des structures dans les 500 systèmes émergents a été comparée à la distribution des structures des systèmes humains répertoriés par la base de donnée UPSID qui contient 451 systèmes (Maddieson, 1984). La figure 12 montre les résultats. Nous observons que les prédictions sont plutôt bonnes, en particulier en ce qui concerne la prédiction des systèmes les plus fréquents pour chaque taille de système de voyelles inférieure à 8. La figure 10 montre une instance du système le plus fréquent chez les humains et dans les simulations. Malgré la prédiction d'un système à 4 voyelles et d'un autre à 5 voyelles qui sont fréquents dans les systèmes émergents (9.1 et 6

pourcent des systèmes) et qui n'apparaissent jamais dans les langues humaines d'UPSID, ces résultats sont d'une qualité comparable à ceux présentés par de Boer (de Boer, 2001). En particulier, nous obtenons toute cette diversité de systèmes et les distributions appropriées en utilisant le même jeu de paramètres, alors que dans les simulations de de Boer, les paramètres varient pour obtenir des systèmes de tailles différentes. Cependant, comme de Boer, le système n'est pas capable de prédire la formation de systèmes à plus de 9 voyelles, qui sont rares mais existent dans les langues humaines. C'est une limite de notre système qui volontairement n'a pas introduit de pression fonctionnelle pour le développement de systèmes efficaces pour la communication dans lesquelles les vocalisations doivent être distinctes les unes des autres.

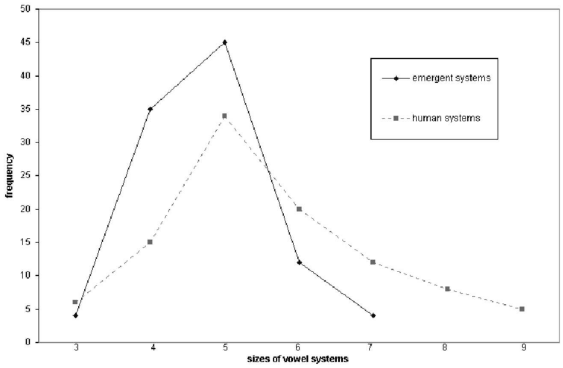

Figure 11 : Distribution des tailles des systèmes de voyelles dans les simulations (trait plein) et dans les langues humaines d'UPSID (trait pointillé).

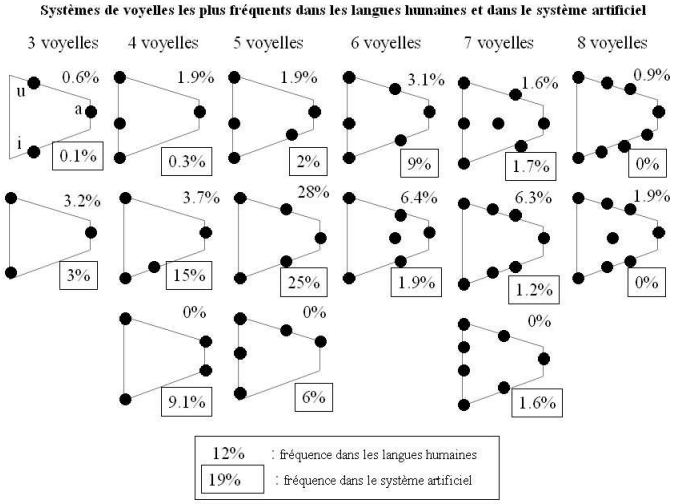

Figure 12 : Comparaison entre la distribution des systèmes de voyelles apparaissant dans le système artificiel et dans les langues humaines (d'après la base de donnée UPSID (Maddieson, 1984))

Conclusion

Nous avons présenté un mécanisme qui fournit une explication possible de la manière dont un code de la parole peut se former dans une communauté d'agents qui ne possèdent pas déjà les moyens de communiquer ou de se coordonner de manière protocolaire (par opposition aux simulations présentées par exemple dans (Kaplan, 2001; de Boer, 2001; Oudeyer, 2001), et qui ne possèdent pas non plus de convention pré-établie ou de systèmes cognitifs spécialisés dans le traitement linguistique (par opposition aux simulations présentées par exemple dans (Kirby, 2001)). Les agents du système n'ont en fait aucune capacité sociale. Nous pensons que la valeur du mécanisme que l'on a présenté réside dans sa qualité d'exemple du type de processus qui a pu permettre au langage, et à la parole en particulier, de se bootstrapper. Nous avons montré comment un pré-requis crucial à la communication, l'existence d'un support physique organisé et conventionnel qui peut véhiculer de l'information entre les membres d'une communauté linguistique, a pu apparaître sans que le langage ne soit déjà là.

Le mécanisme auto-organisé du système apparaît aussi comme un complément nécessaire à l'explication néo-Darwinienne classique des origines de la parole. Il est compatible avec le scénario néo-Darwinien classique dans lequel l'environnement favorise la réplication des individus qui sont capables de parole. Dans ce scénario, notre système artificiel joue le même rôle que les lois de la physique des gouttes d'eau dans l'explication de l'origine des formes hexagonales des cellules de cire construites par les abeilles. Nous avons montré comment des mécanismes auto-organisés peuvent faciliter le travail de la sélection naturelle en contraignant l'espace des formes biologiques. En effet, notre système montre que la sélection naturelle n'a pas eu nécessairement besoin de trouver des génomes qui pré-programmaient le cerveau de manière précise et spécifique de manière à créer et à apprendre des codes de la parole digitaux. La capacité de coordination sociale protocolaire et le comportement d'imitation explicite sont aussi des exemples de mécanismes qui ne sont pas nécessaires pour la formation d'un premier code de la parole, comme le démontre notre système. Cela permet de tracer les contours d'un scénario néo-Darwinien convaincant, en éclairant les zones d'ombre conceptuelles qui en faisaient une idée vague plutôt qu'un mécanisme opérationnel.

En outre, ce même mécanisme permet de rendre compte de plusieurs propriétés de la parole : digitalité, combinatorialité, tendances universelles et diversité. Nous pensons que cette explication est originale parce que : 1) un seul mécanisme est utilisé pour rendre compte de toutes ces propriétés; 2) nous n'utilisons pas de pression explicite pour développer un système de communication efficace, et nous n'utilisons pas de systèmes cognitifs qui soit spécifique à la parole (les mêmes structures neurales pourraient être utilisées pour apprendre la coordination main-œil par exemple). En particulier, en ayant fait des simulations à la fois avec et sans non-linéarités dans la fonction qui fait correspondre des sons à des configurations articulatoires, nous avons pu montrer qu'en principe, alors que la fréquence de tel ou tel phonème est influencée par ces non-linéarités, leur existence même – ç'est à dire le codage phonémique de la parole – ne requiert pas la présence de non-linéarités, mais peut être un effet de bord de la dynamique du couplage des cartes neurales sensori-motrices. Cela contraste avec certaines théories existantes qui considèrent que le codage phonémique s'explique soit par la présence de non-linéarités (Stevens, 1972; Mrayati *et al.,* 1988), soit par la présence d'une pression explicite pour maximiser la distinctivité perceptuelle entre les vocalisations (Lindblöm, 1992).

Un modèle comme celui de de Boer (de Boer, 2001) peut être considéré comme s'appliquant à un phénomène plus récent dans l'histoire évolutionnaire de la parole. Plus précisément, le modèle de de Boer, ainsi que par exemple celui décrit dans (Oudeyer, 2001) sur la formation des systèmes de syllabes, s'intéressent au recrutement de codes de la parole comme ceux qui sont apparus dans nos simulations, et étudient comment ils sont modelés sous l'influence d'une pression fonctionnelle pour la communication. En effet, alors que nous avons montré ici que l'on pouvait expliquer un certain nombre de phénomènes de la parole sans utiliser une telle pression, certaines autres propriétés de la parole ne peuvent être expliquées qu'en y faisant appel. C'est le cas par exemple de l'existence des grands systèmes de voyelles (Schwartz *et al.,* 1997).

Annexe

Les détails techniques du système artificiel : les neurones n_i ont une fonction d'activation gaussienne. Si on note $G_{i,t}(s)$ la fonction d'activation de n_i au temps t, s un stimulus, v_i le vecteur préféré (les poids) de n_i, alors la fonction d'activation est :

$$G_{i,t}(s)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{1}{2}v_i.s^2/\sigma^2}$$

où $v_1.v_2$ dénote le produit scalaire entre les deux vecteurs v_1 et v_2. Le paramètre σ^2 détermine la largeur des gaussiennes, et s'il est grand les gaussiennes sont larges (une valeur de 0.05 comme on utilise dans les simulations signifie que le neurone associé s'activera significativement pour environ 10 pourcent de l'espace des stimuli).

Quand un neurone dans une carte perceptuelle est activé par un stimulus *s*, alors son vecteur préféré est mis à jour :

$$v_{i,t+1}=v_{i,t}+0.001.G_{i,t}(s).(s-v_{i,t})$$

Quand un son est perçu et provoque une activation qui est propagée jusqu'aux neurones moteurs, les vecteurs préférés des neurones moteurs sont aussi modifiés. Le vecteur préféré du neurone moteur le plus actif est sélectionné et utilisé comme référence : les autres vecteurs préférés sont modifiés de manière à ce qu'ils se rapprochent de la référence. Cette modification est réalisée avec exactement la même formule que pour les neurones perceptuels, avec *s* qui est maintenant le vecteur préféré référence. Quand un agent entend un vocalisation qui a été produite par lui-même, et donc que ses neurones moteurs sont déjà activés quand le son active les neurones perceptuels, alors les poids des connections entre les neurones perceptuels et les neurones moteurs sont modifiés. Une règle d'apprentissage hebbienne est mise en œuvre. Si n_i est un neurone de la carte perceptuelle connecté à n_j qui est un neurone de la carte motrice, alors le poids $w_{i,j}$ qui caractérise leur connexion change :

$$\Delta w_{i,j}(t)=c_2.(G_{i,s_i}(t)-\langle G_{i,s_i}(t)\rangle).(G_{j,s_j}(t)-\langle G_{j,s_j}(t)\rangle)$$

où s_i et s_j sont les entrées des neurones n_i et n_j , $\prec G_{i,s_i}(t)\succ$ la moyenne des activations du neurone n_i dans le passé récent, et c_2 une petite constante.

En outre, les neurones de chaque carte son totalement interconnectés. Ces connections ont des poids qui sont symétriques. Cela permet d'avoir une dynamique similaire aux réseaux de neurones de Hopfield, avec des points attracteurs qui sont utilisés pour modéliser le comportement de catégorisation. Les poids représentent la corrélation des activités entre les neurones, et sont appris avec la même loi hebbienne :

$$\Delta w_{i,j}(t)=c_2.(G_{i,s_i}(t)-\prec G_{i,s_i}(t)\succ).(G_{j,s_j}(t)-\prec G_{j,s_j}(t)\succ)$$

Ces connections sont utilisées pour la relaxation de chaque carte neurale, après que les activations initiales de la carte perceptuelle ont été propagée à la carte motrice et utilisées pour mettre à jour les vecteurs préférés et les connections entre les cartes. La relaxation est une mise à jour des niveaux d'activation de chaque neurone en utilisant la formule :

$$act(i,t+1)=\frac{\sum_j w_{i,j}.act(j,t)}{\sum_j act(j,t)}$$

où $act(i,t)$ est l'activation du neurone n_i au temps *t*.

Pour visualiser l'évolution des activations de tous les neurones pendant la relaxation, nous utilisons le « vecteur population ». En effet, l'activation de tous les neurones dans la carte peut être résumée par ce vecteur population développé par Georgopoulos (Georgopoulos *et al.,* 1988) : c'est la somme de tous les vecteurs préférés des neurones pondérés par le niveau d'activité des neurones qui leurs sont associés (et normalisée) :

$$pop(v)=\frac{\sum_j act(j).v_i}{\sum_j act(j)}$$

Références bibliographiques

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

D'Arcy Thompson (1961). *On Growth and Form*. Cambridge University Press. (texte original : 1932).

Batali, J. (1998). 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: Cambridge University Press.

de Boer B. (2001). *The origins of vowel systems*. Oxford Linguistics, Oxford University Press.

Browman C., Goldstein L. (1986). Towards an articulatory phonology. Phonology Yearbook 3, 219–252.

Cangelosi A., Parisi, D. (2002). *Simulating the Evolution of Language*. London : Springer.

Crothers J. (1978). Typology and universals of vowels systems. *Phonology*, Vol. 2, 93-152.

Georgopoulos A.P., Kettner R.E., Schwartz A.B. (1988). Primate motor cortex and free arm movement to visual targets in three-dimensional space. II. Coding of the direction of movement by a neuronal population. *Journal of Neurosciences*, Vol. 8, 2928-2937.

Hurford J., Studdert-Kennedy M., Knight C. (1998*)*. *Approaches to the evolution of language*. Cambridge : Cambridge University Press.

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

Kauffman S. (1996). *At Home in the Universe : The Search for Laws of Self-Organization and Complexity*. Oxford University Press.

Kandel E.R., Schwartz J.H., Jessell T.M. (2001). *Principles of Neural Science*, McGraw-Hill/Appleton and Lange.

Kirby S. (1998) Fitness and the selective adaptation of language. In Hurford, J. R., Studdert-Kennedy, M. and Knight C., editors, *Approaches to the Evolution of Language: Social and Cognitive Bases*. Cambridge: Cambridge University Press.

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

Ladefoged P., Maddieson I. (1996). *The Sounds of the World's Languages*. Oxford : Blackwell Publishers.

Lindblöm B. (1992). Phonological Units as Adaptive Emergents of Lexical Development. in Ferguson, Menn, Stoel-Gammon (Eds.) *Phonological Development : Models, Research, Implications*. York Press, Timonnium, MD, 565-604.

Maddieson I. (1984). *Patterns of Sounds*. Cambridge : Cambridge University Press.

Mrayati M., Carr R., Gurin B. (1988). Distinctive region and modes : A new theory of speech production. *Speech Communication,* Vol. 7, 257–286.

Nicolis G., Prigogine I. (1977). *Self-Organization in Nonequilibrium Systems : From Dissipative Structures to Order through Fluctuations*. Wiley.

Oudeyer P-Y. (2001). Origins and Learnability of Syllable Systems, a Cultural Evolutionary Model. In *Artificial Evolution*, Collet P., Fonlupt C., Hao J.K., Lutton E., Schonenauer M., (Eds). LNCS 2310, Springer Verlag, 143-155.

Oudeyer P-Y. (2003). *L'auto-organisation de la parole*. Thèse de Doctorat de l'Université Paris VI, www.csl.sony.fr/~py/theseFrench.html

Oudeyer P-Y. (2005). The Self-Organization of Speech Sounds. *Journal of Theoretical Biology*, 233(3), 435-449.

Schwartz J.L., Boé L.J., Vallée N., Abry C. (1997). Major trends in vowel systems inventories. *Journal of Phonetics*, Vol. 25, 233-253.

Steels L. (1997). The synthetic modeling of language origins. *Evolution of Communication*, 1(1), 1-35.

Stevens K., (1972). *The quantal nature of speech: evidence from articulatoryacoustic data*. New-York : Mc Graw-Hill, 51–66.

Studdert-Kennedy, M. and Goldstein, L. (2003) Launching language: The gestural origin of discrete infinity. In M.H. Christiansen and S. Kirby, editors, *Language Evolution: The States of the Art*. Oxford University Press.

Pour approfondir

Pour une vision d'ensemble de la recherche sur les origines du langage, le recueil d'article (Hurford *et al.,* 1998) est un bon point de départ. Pour découvrir en détail d'autres exemples de modèles computationnels des origines du langage et de la parole, (Cangelosi et Parisi, 2002) est une référence standard. Pour avoir une vision générale de la recherche sur les origines de la parole et sur les liens entre « parole » et « auto-organisation », (Lindblöm, 1992), (de Boer, 2001), et (Studdert-Kennedy and Goldstein, 2003) sont des classiques. Pour plus de détails sur le travail présenté dans ce dossier, voir (Oudeyer, 2003; Oudeyer, 2005).

De manière plus générale, pour comprendre le rôle de l'auto-organisation dans l'origine des formes dans le monde biologique, et en particulier de la relation avec le concept de sélection naturelle, (Kauffman, 1996) est un bon point de départ. Ma thèse y consacre aussi un chapitre que j'ai tenté de rendre didactique (Oudeyer, 2003). Ensuite, (D'Arcy Thompson, 1932) et (Ball, 2001) sont des mines inépuisables d'exemples concrets et détaillés.

L'auteur

Pierre-Yves Oudeyer a fait des études d'informatique à l'Ecole Normale Supérieure de Lyon, un DEA d'intelligence artificielle à Paris VI, et une thèse au Sony Computer Science Laboratory à Paris dans laquelle il a développé une théorie computationelle des origines de la parole. Il est actuellement chercheur à Sony CSL Paris. Il s'intéresse aux origines de la parole, et construit des sociétés de robots pour étudier les mécanismes qui permettent à des codes linguistiques de se former culturellement et grâce à des phénomènes d'auto-organisation. Il conduit aussi des recherches dans le domaine

de la robotique développementale, et étudie comment des robots peuvent créer de manière autonome leurs propres buts de manière à ce que la complexité de leurs activités et de leurs savoir-faire croissent de manière continue et ouverte, sans qu'il n'y ait de supervision. En particulier, il travaille sur des algorithmes de curiosité adaptative.