



STEVAN HARNAD
HARNAD@ECS.SOTON.AC.UK
COGNITIVE SCIENCES CENTRE
UNIVERSITY OF SOUTHAMPTON
[HTTP://WWW.COQSCI.SOTON.AC.UK/~HARNAD/](http://www.cogsci.soton.ac.uk/~harnad/)

ANGELO CANGELOSI
CENTRE FOR NEURAL AND ADAPTIVE SYSTEMS
UNIVERSITY OF SOUTHAMPTON

TRADUCTION:
MICHÈLE COULMANCE
MICHELE.COULMANCE@EPFL.CH
SERVICE INFORMATIQUE CENTRAL EPFL



LE MARTIEN, LES CHAMPIGNONS ET LES RÉSEAUX DE NEURONES

RÉSUMÉ

En utilisant des réseaux de neurones formels pour simuler l'apprentissage, il s'agit de mettre en compétition directe deux manières d'acquérir des connaissances fort utiles au demeurant, puisqu'il s'agira d'acquérir un comportement adapté face à un champignon vénéneux ou comestible.

Une première méthode est l'apprentissage en temps réel par essai-et-erreur, basé sur les entrées sensorielles, visuelles par exemple, les conséquences du succès du tri fournissant un *feed-back* immédiat.

Une seconde méthode consiste à utiliser des entrées bien différentes, basée cette fois sur l'*oui-dire* elle permet d'apprendre à catégoriser grâce à l'utilisation d'étiquettes symboliques combinées.

Comparer ainsi les résultats de ces deux méthodes d'acquisition peut contribuer à cerner la nature de l'avantage adaptatif du langage oral.

LANGAGE ET ÉVOLUTION: LE POINT DE VUE DU MARTIEN

Si l'on s'en réfère aux Darwiniens, en matière d'évolution tout est affaire de survie et de reproduction. Et bien des espèces perdurent sans avoir jamais prononcé un mot.

Alors qu'est-ce que cette capacité langagière, telle que les humains possèdent aujourd'hui un organe *dédié au langage*, le cerveau (tout au moins certaines aires spécialisées de ce cerveau), et un penchant à passer une bonne partie de leur existence à utiliser cette capacité ?

Que resterait-il de notre espèce aujourd'hui si nos aptitudes langagières disparaissaient ?

Quel est donc l'aspect critique du langage oral qui nous a conduit inexorablement sur ce chemin de l'évolution, chemin unique se distinguant de tous ceux empruntés par les espèces dépourvues de langage ?

Les conjectures ne manquent pas sur les apports du langage d'un point de vue évolutif: le langage a augmenté nos performances dans le domaine de la chasse; il nous a aidé à construire des outils; il nous a ouvert la voie de la socialisation. Mais comment tester ces hypothèses ? Le silence du langage est criant, tant en archéologie qu'en paléontologie !

À la manière d'un exobiologiste, peut-être pourrions-nous essayer d'adopter le célèbre point de vue de l'anthropologue Martien: les humains ont été capables d'agir efficacement dans leur monde, et s'ils en ont été capables, c'est qu'ils possèdent des connaissances à propos de ce monde. Et le Martien pourrait ainsi décrire ces connaissances: quels genres d'objets existe-t-il dans ce monde, et quel genre d'actions peut-on exercer sur eux ? En d'autres termes, la base de connaissances est une base de connaissances catégorielle.

Il pourrait ensuite s'interroger sur la provenance de ces connaissances: pour certaines elles semblent innées; d'autres semblent acquises sur la base d'un apprentissage individuel par essai-et-erreur; mais notre Martien sera surpris en constatant que la plupart de ces catégories sont élaborées non pas sur la base d'un apprentissage individuel par essai-et-erreur, mais bien plutôt sur la seule base d'un bref échange vocal, *STOP! vénéneux!*, avec un individu qui lui-même a eu un bref échange vocal avec... Et ce mode là de capture de connaissances, sans le dur apprentissage de l'expérience directe, est en quelque sorte une capture non plus laborieuse, mais frauduleuse! délit sans victime cependant.

Le mécanisme par lequel peut s'opérer cette capture est simple à décrire: tout commence par la constitution, par essai-et-erreur, d'un répertoire de catégories; cet apprentissage repose sur un mécanisme nous permettant d'extraire des traits caractéristiques communs à partir d'entrées sensori-motrices (les célèbres cinq sens plus la motricité).

A ces catégories de base on associe un nom symbolique arbitraire, et il s'agit bien d'un signe arbitraire, ceci est familier à notre Martien, qui connaît les éternelles vérités platoniciennes logiques et mathématiques, et qui sait qu'elles peuvent toujours être codées sous forme de notation symbolique.

Les épisodes vocaux échangés peuvent être décrits comme des propositions, des combinaisons de signes-symboles qui peuvent être interprétées comme *juste ou faux*: ceci est un champignon, il a des points noirs: ce champignon est comestible/vénéneux ? Comestible !

Le Martien comprendra vite que ces épisodes vocaux concernent l'appartenance d'un objet à telle ou telle catégorie, et que la combinatoire rend possible l'acquisition de nouvelles catégories en sautant l'étape combien dangereuse (s'il s'agit de champignons !) de l'apprentissage par essai-et-erreur.

Ce mécanisme d'acquisition, spécifiquement permis par le langage, est-il un des avantages adaptatifs du point de vue de l'évolution de l'espèce ? Et comment tester cette hypothèse ?

Justement en utilisant une simulation informatique: imaginons un monde dans lequel poussent différentes espèces de champignons, monde peuplé de petites créatures qui doivent être capables de reconnaître les champignons comestibles des vénéreux pour survivre et se reproduire.

Quelques considérations théoriques s'imposent avant d'entrer dans ce monde virtuel.

FONDATION DU SIGNE

L'utilisation des symboles –surtout les noms de catégories– va permettre, par héritage de propriétés inhérentes et par combinaison à d'autres symboles, la construction de nouveaux symboles: par exemple **zèbre**, qui hérite des propriétés de **cheval** et de **rayé**. Et ce symbole, **zèbre**, pourra être le vecteur de transmission d'une connaissance à propos d'une nouvelle catégorie, alors même qu'il n'y a pas eu d'interaction sensori-motrice directe avec ce nouvel objet. Ce qui peut présenter un intérêt certain dans le monde des champignons: qui ne souhaiterait pouvoir éviter les essais et erreurs pour apprendre quelle espèce de champignons n'est pas comestible ?

Mais cette construction doit avoir des fondations solides. En dernière analyse il faut bien qu'on puisse associer un symbole avec **quelque chose** – ou **quelques choses** – qu'on a directement échantillonné de manière sensori-motrice; de même que la valeur d'une monnaie doit finalement s'appuyer sur un étalon or, ou une autre ressource ayant une existence matérielle. Le rôle de la perception catégorielle est certainement essentiel pour que puissent se construire les premières associations entre un genre (ou catégorie) d'objet et le nom associé à ce genre d'objet.

PERCEPTION CATÉGORIELLE

Nous possédons certainement dès la naissance des *détecteurs de catégories*, la grenouille naît avec la capacité de détecter une mouche; sans nécessiter d'essai-et-erreur, la simple exposition à ce stimulus déclenchera la reconnaissance, et la capture. Au-delà de ces capacités innées, nos connaissances résultent en partie de nos interactions avec le monde: apprentissage en temps réel, par essai-et-erreur, le *feed-back* étant fourni par les conséquences d'un tri correct ou incorrect. Puisque dans bien des cas le tri *correct* est loin d'être trivial, notre Martien pourrait noter que nous devons posséder des *périphériques sensoriels performants* (organes sensoriels) qui nous permettent d'extraire certaines caractéristiques critiques des objets; et de les classer ensuite sur cette base dans des catégories adéquates, sur lesquelles nous pourrions appliquer des actions spécifiques pertinentes. Ces catégories doivent apparaître comme des catégories discrètes, sans recouvrement possible avec d'autres.

On appelle perception catégorielle le phénomène qui va conduire à une perception discrétisée de stimuli pouvant varier de manière continue: la perception des couleurs par exemple, la manière dont le spectre va être artificiellement subdivisé en *violet-indigo-bleu-vert-jaune-orangé-rouge*. Il va y avoir distorsion au niveau de nos représentations internes, diminution des différences perçues entre les exemplaires classés dans une même catégorie, augmentation des différences perçues entre des exemplaires de catégories différentes (moindre écart perçu entre deux *sortes* de vert qu'entre vert et jaune, alors que l'écart en terme physique est le même). Cette opération de **compression/dilatation** permet de placer des limites tout ou rien entre des régions de **l'espace de similitude** occupées par des objets qui vont recevoir des attributs différents.

Le phénomène de perception catégorielle est observé tant avec les catégories innées qu'avec les catégories acquises par apprentissage, chez les humains comme chez les animaux. Il apparaît également dans les **réseaux de neurones** formels avec apprentissage par *feed-back* rétroactif (rétropropagation de l'information sur l'écart entre réponse fournie et objectif à atteindre).

La mesure de cette compression/dilatation dans l'espace de similitude peut être prise comme un indice d'une catégorisation efficace.

LE MONDE DES CHAMPIGNONS

La simulation informatique va se dérouler dans un monde de champignons où naissent, vivent et meurent de petits individus fouineurs qui vont apprendre ce qu'il faut faire avec ces champignons, les manger, les **marquer**, éventuellement revenir à leur emplacement. Notons que ces individus virtuels, au nombre de 100, vont pouvoir donner naissance à de nouvelles générations.

Les champignons ayant la caractéristique A (par ex. avec des points orange sur le chapeau) peuvent être mangés. Les champignons ayant la caractéristique B (par ex. avec un pied grisé) doivent être **marqués**, les champignons ayant les deux caractéristiques A et B doivent être marqués et mangés. Et les petits individus virtuels devront apprendre à ignorer les autres caractéristiques C, D, E qui peuvent être présentes ou non.

On peut donner une signification écologique au comportement de **marquage** en supposant que les champignons B produisent une toxine, mais qu'un comportement quelconque, recouvrir ces champignons par ex. permet de bloquer les effets toxiques. De surcroît, les seuls champignons AB ont une particularité supplémentaire: à l'emplacement sur lequel ils poussent, beaucoup d'autres champignons de la même espèce pousseront. Il est donc intéressant d'apprendre qu'on doit revenir à ces emplacements particuliers.

Le petit individu virtuel fouineur a la capacité de se déplacer, (quatre possibilités, tourner de 90° à gauche, tourner de 90° à droite, faire un pas en avant, rester sur place) et pour guider ses déplacements, il va recevoir des indications concernant uniquement le champignon le plus proche de lui.

Il a aussi une capacité de vocalisation: lorsqu'il s'approche d'un champignon, il émet un signal associé avec l'action qu'il se propose d'appliquer au champignon (manger, marquer).

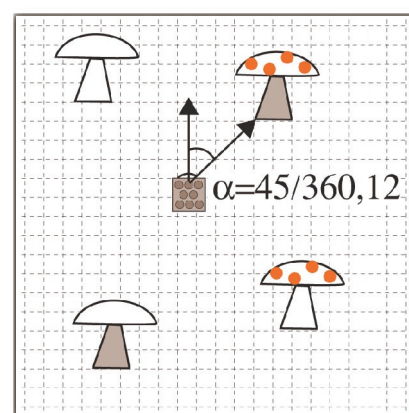


ILLUSTRATION DU MONDE VIRTUEL EN DEUX DIMENSIONS AVEC UN INDIVIDU ET QUATRE SORTES DE CHAMPIGNONS.

LES ENTRÉES REÇUES CONCERNENT

- L'absence ou la présence (1 ou 0) des caractéristiques A,B,C,D,E. Un champignon AB sera donc représenté comme 11***, * étant 1 ou 0 pour les caractéristiques C,D,E, qui peuvent être présentes ou absentes, mais sont irrelevantes.
- La localisation du champignon par rapport à l'individu, exprimée sous forme d'angulation normalisée.

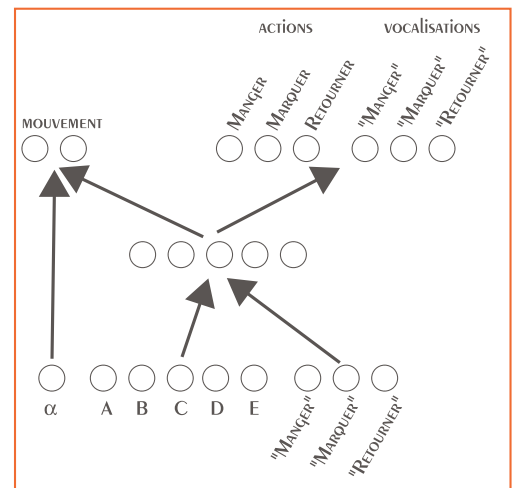
Les petits fouineurs devront apprendre à catégoriser les champignons sur la base des informations reçues, et fournir les réponses adaptées:

- Manger
- Marquer.

Ils ont la capacité d'émettre des vocalisations associées à ces comportements. Et ils peuvent parfois **recevoir en entrée les vocalisations émises par un autre individu**. Ceci afin de tester le rôle adaptatif que peut avoir la capture de connaissances par ouï-dire. Le réseau de neurones possède une couche cachée de cinq unités, le poids des connexions entre les différentes unités des différentes couches (entrée, sortie, cachée) est distribué au hasard.

Au cours d'un cycle de vie un individu est confronté à différentes distributions de 40 champignons, et un cycle est constitué de 2000 *actions*. Une action consiste en deux propagations d'activation dans le réseau de neurones:

- L'individu produit d'abord un mouvement, et une action ainsi que la vocalisation associée, ceci sur la base des caractéristiques ABCDE fournies en entrée. Il y a apprentissage supervisé basé sur l'algorithme de rétro-propagation. Le superviseur étant capable de calculer l'écart entre la réponse fournie et l'objectif à atteindre, un algorithme d'ajustement des poids synaptiques va permettre de diminuer l'efficacité de certaines connexions et de renforcer l'efficacité de certaines autres de telle sorte que le comportement fourni lors d'une autre session soit plus proche du comportement à atteindre.
- Une seconde phase de propagation d'activité va permettre à l'individu d'apprendre également à imiter les vocalisations. Sur la base de l'entrée uniquement vocalisation cette fois, il apprend selon le même principe de rétro-propagation de gradient à émettre en sortie la vocalisation correcte.



SCHEMA DE L'ARCHITECTURE DU RESEAU DE NEURONES

Les individus, divisés en deux groupes, vont connaître deux stades d'existence, comportant 2000 actions pour chaque stade.

Lors du premier stade, ils apprennent tous par essai-erreur à associer le comportement **mange** à la caractéristique A, le comportement **marque** à la caractéristique B des champignons auxquels ils sont confrontés. Les champignons AB sont mangés et marqués.

Dans un second stade, il s'agira d'apprendre à associer le comportement dit de retour aux champignons AB:

Capture sensori-motrice: Un groupe va apprendre par essai-erreur sur la base des caractéristiques AB, avec supervision et rétropropagation.

Capture symbolique: L'autre groupe va apprendre, également avec supervision, mais cette fois uniquement sur la base d'une entrée **vocalisation** (d'une dénomination). Rappelons que les vocalisations sont associées au comportement appliqué au champignon.

Conditions		ENTRÉES		RÉTRO-PROPAGATION	
		CARACTÉRISTIQUES	VOCALISATION	COMPORTEMENT	VOCALISATION
CAPTURE SENSORI-MOTRICE	MANGER-MARQUER	OUI	NON	OUI	OUI
	RETOURNER	OUI	NON	OUI	OUI
CAPTURE SYMBOLIQUE	RETOURNER	NON	OUI	OUI	OUI
IMITATION	IMITATION	NON	OUI	NON	OUI

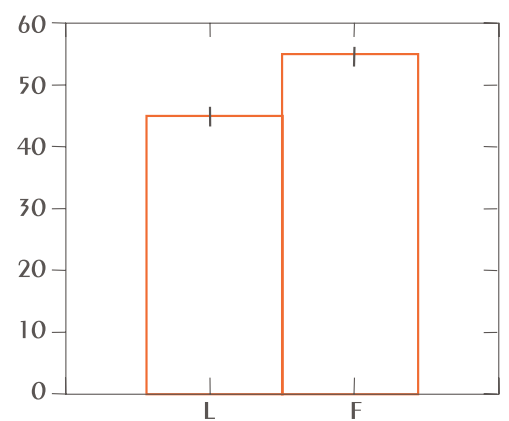
On peut dans un premier temps comparer les performances comportementales des deux groupes en comptabilisant le nombre de retours vers les champignons AB.

Le mécanisme d'acquisition par oui-dire symbolique, rendu possible par le langage, semble donc bien avantageux.

On peut alors se demander ce qui a changé au cours de ces apprentissages, et en particulier observer l'activité des cinq unités de la couche cachée du réseau de neurones.

Nous n'entrerons pas dans le détail des traitements effectués, les calculs appliqués à ces activités ont pour objet la mesure des distances intra et intercatégorielles dans l'espace de similitude, ce qui, rappelons-le peut être un indice d'une catégorisation efficace.

Il est à noter que la distance la plus grande entre **retour** et les autres catégories apparaît lorsque la catégorisation a été acquise sur la base de la capture symbolique.



NOMBRE DE RETOURS VERS LES CHAMPIGNONS AB POUR LES GROUPES L (CAPTURE SENSORI-MOTRICE) ET F (CAPTURE PAR OUI-DIRE)

La stratégie d'acquisition par capture symbolique se révèle donc bien plus efficace que la stratégie d'acquisition sensori-motrice; on observe toujours une déformation de l'espace de similitude, avec compression des écarts entre membres d'une catégorie et dilatation des écarts entre les catégories différentes, et cette déformation se produit dès qu'il y a apprentissage sensori-moteur par essai-erreur; l'apprentissage symbolique va permettre non seulement d'hériter de cette déformation, mais encore de l'accroître, rendant ainsi la catégorisation plus efficace au niveau comportemental.

Conclusions

Le langage a une influence sur la manière dont nous percevons le monde; pour notre Martien anthropologue, cette influence peut approximativement se concevoir ainsi: toutes les espèces de cette planète apprennent à catégoriser les objets et les événements auxquels ils sont confrontés; l'acquisition de cette aptitude à catégoriser se fait à l'échelle de l'existence individuelle, et à l'échelle plus générale, cumulative, des différentes générations.

Confrontés au monde extérieur, les individus ayant construit ces représentations catégorielles pourront apprendre quel est le comportement le plus adapté dans telle ou telle situation.

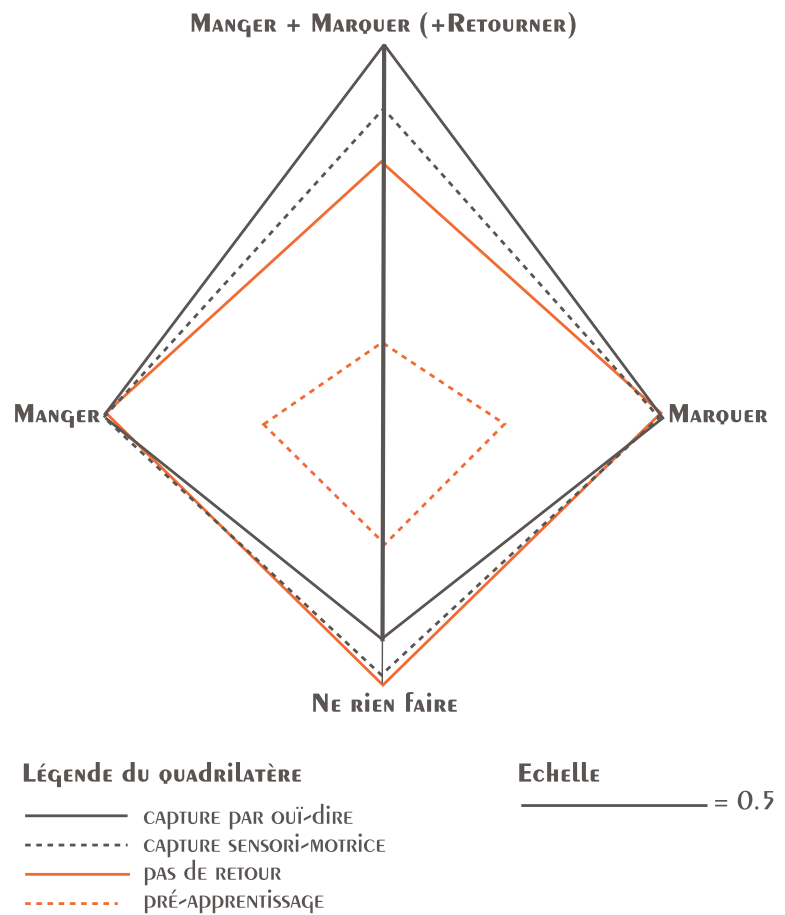
Ce faisant, ils apprennent à voir le monde différemment, à détecter des invariants, à minimiser les écarts entre traits similaires, à augmenter les écarts entre traits distinctifs; le tout guidé par le *feed-back* constitué par les conséquences avantageuses ou désavantageuses de leur tri. Mais une seule espèce a découvert un autre mode d'apprentissage: assigner un nom symbolique aux catégories, ce qui a donné accès à un monde de possibilités combinatoires où s'instaure un véritable courant d'échanges entre catégories.

Les échanges entre catégories comme ceux qui se font entre les combinaisons de symboles conduisent aussi à un véritable brassage des points de vue. Tout ce *commerce* a comme conséquence un réarrangement des représentations internes des catégories, se manifestant tantôt par une subtile compression résultant de l'apprentissage que X est membre de Y, tantôt par une restructuration plus importante résultant de quelque découverte scientifique plus radicale.

Seul notre Martien connaît les conditions spécifiques initiales dans lesquelles le pouvoir génératif de l'attribution de noms, ainsi que leur combinatoire booléenne, ont imposé leurs effets biologiques sur notre planète.

Mais notre simulation permet peut-être d'envisager comment les bénéfices de ce processus se sont multipliés, comme des champignons, conduisant nos ancêtres, forts de leurs réussites, à nommer les catégories, et à établir des liens entre les noms pour décrire de nouvelles catégories, avec de plus en plus de zèle !

Il est cependant légitime de se demander si un monde-jouet virtuel à ce point simplifié peut vraiment nous apporter quelques lumières sur un phénomène aussi riche et complexe que l'importance adaptative du langage. Ce genre de résultat peut-il vraiment être transposé dans le monde réel ? Ce problème de validité du modèle est d'ailleurs assez général dans bien des domaines de la modélisation cognitive qui souvent ne propose pas de tâches à l'échelle humaine.



PROJECTION 2D DES DISTANCES ENTRE CATEGORIES DANS LES DIFFÉRENTES CONDITIONS

Pour le moment nous pouvons seulement essayer d'enrichir ce modèle, en y injectant davantage de la complexité et des contraintes du monde réel.

D'APRÈS:

Cangelosi, A. & Harnad, S. (2001) *The Adaptive Advantage of Symbolic Theft Over Sensorimotor Toil: Grounding Language in Perceptual Categories*. *Evolution of Communication* 4 (1). 117-141.
<http://cogprints.soton.ac.uk/documents/disk0/00/00/20/36/index.html>

QUELQUES RÉFÉRENCES:

- ▶ **Cangelosi A., Greco A. & Harnad S.** (2002) *Symbol Grounding and the Symbolic Theft Hypothesis*. In: Cangelosi, A. & Parisi, D. (Eds.) – *Simulating the Evolution of Language*. London, Springer.
<http://cogprints.soton.ac.uk/documents/disk0/00/00/21/32/index.html>
- ▶ **Harnad S.** (1987) *Category Induction and Representation*, In: Harnad, S. (ed.) (1987) *Categorical Perception: The Groundwork of Cognition*. New York: Cambridge University Press.
<http://cogprints.soton.ac.uk/documents/disk0/00/00/15/72/index.html>
- ▶ **Harnad S.** (1990) *The Symbol Grounding Problem*. *Physica D* 42: 335-346.
<http://cogprints.soton.ac.uk/documents/disk0/00/00/06/15/index.html>
- ▶ **Harnad S.** (2002) *Symbol Grounding and the Origin of Language*. In: M. Scheutz (ed.) *Computationalism: New Directions*. Cambridge MA: MIT Press. Pp. 143-158.
<http://cogprints.soton.ac.uk/documents/disk0/00/00/21/33/index.html>
- ▶ **Harnad S.** (2003) *Categorical Perception*. *Encyclopedia of Cognitive Science*. Nature Publishing Group. Macmillan.
<http://www.ecs.soton.ac.uk/~harnad/Temp/catperc.html>
- ▶ **Harnad S.** (2003) *Symbol-Grounding Problem*. *Encyclopedia of Cognitive Science*. Nature Publishing Group. Macmillan. <http://www.ecs.soton.ac.uk/~harnad/Temp/symgro.html>. ■

