

Peter Gärdenfors*

Les Espaces Conceptuels*

En sciences cognitives, il existe actuellement deux approches majeures au problème de la modélisation des représentations. L'approche *symbolique* est basée sur l'hypothèse selon laquelle la cognition consiste essentiellement en la manipulation de symboles. La deuxième approche est le connexionisme, qui modélisent des associations par des réseaux de neurones artificiels. Dans cet article, je propose une troisième façon de représenter des informations qui est basée sur des structures *géométriques* plutôt que des symboles ou des connexions entre neurones. En employant ces structures, des relations de *similarité* peuvent être modélisées de façon naturelle. Une théorie des *espaces conceptuels* est esquissée. Un tel espace est construit à partir de représentations géométriques basées sur un certain nombre de *dimensions de qualité* qui sont souvent dérivées de mécanismes perceptives. L'article indique comment des espaces conceptuels peuvent représenter diverses sortes d'information, et comment ils peuvent être utilisés afin de décrire des processus cognitifs comme la formation de concepts et l'induction. Un modèle "géométrique" de la formation des concepts est proposé et, entre autres, sa relation à la théorie des prototypes est discutée.

Mots clés: Espaces conceptuels, modèles géométriques, dimensions de qualités, similarité, représentations symboliques, connexionisme, formation des concepts, théorie des prototypes.

Conceptual spaces. Within cognitive science, there are currently two dominating approaches to the problem of modeling representations. The *symbolic* approach starts from the assumption that cognition is essentially involving symbol manipulation. The second approach is connectionism which models associations by artificial neuron networks. In this article, I advocate a third form of representing information that is based on using *geometrical* structures rather than symbols or connections between neurons. Using these structures *similarity* relations can be modelled in a natural way. A theory of *conceptual spaces* is outlined. Such a space is built up from geometrical representations based on a number of *quality dimensions* which are often derived from perceptual mechanisms. It will be outlined how conceptual spaces can represent various kind of information and how they can be used to describe cognitive processes like concept formation and induction. A "geometric" model of concept

* Lund University Cognitive Science, Kungshuset, S-222 22 Lund, Suède
Peter.Gardenfors@lucs.lu.se

*Cet article constitue un précis de mon livre *Conceptual Spaces*, MIT Press, 2000. Le texte est traduit par Marie-Claude Lorne.

formation will be proposed and, among other things, its relation to prototype theory will be discussed.

Keywords: Conceptual spaces, geometric models, quality dimensions, similarity, symbolic representations, connectionism, concept formation, prototype theory.

1. MODELISER LES REPRESENTATIONS

Les sciences cognitives ont deux buts principaux. Le premier est un but *explicatif* : grâce à l'étude des activités cognitives des êtres humains et des autres animaux, on peut formuler des *théories* portant sur différents aspects de la cognition. On teste ces théories au moyen d'expériences ou de simulations par ordinateur. L'autre but est *constructif* : en construisant des *artefacts* comme des programmes de jeu d'échec, des robots, des animats, etc., on essaie de construire des systèmes qui sont capables d'accomplir diverses tâches cognitives. Ces deux buts se heurtent à un problème clef : comment modéliser de manière appropriée les *représentations* utilisées par le système cognitif ?

Deux types d'approche du problème de la modélisation des représentations dominent actuellement les sciences cognitives. L'approche *symbolique* est fondée sur l'hypothèse qu'on doit modéliser les systèmes cognitifs au moyen de machines de Turing. D'après cette position, la cognition met essentiellement en jeu une manipulation de symboles. La seconde approche est l'*associationnisme*, dans lequel des associations entre différents types d'éléments informationnels remplissent l'essentiel de la tâche représentationnelle. Le connexionnisme est un cas particulier d'associationnisme, où l'on modélise les associations au moyen de réseaux de neurones artificiels. L'approche symbolique et l'approche associationniste présentent toutes deux des avantages et des inconvénients. On les présente souvent comme des paradigmes concurrents, mais puisqu'elles traitent les problèmes cognitifs à des niveaux différents, je soutiendrai qu'on devrait plutôt les considérer comme des méthodologies complémentaires.

Toutefois, ni la représentation symbolique, ni le connexionnisme ne semblent offrir les outils de modélisation appropriés pour rendre compte de certains aspects de la cognition. Je défendrai dans cet article une troisième forme de représentation de l'information, qui est fondée sur l'utilisation de structures *géométriques* plutôt que sur des symboles ou des connexions entre des neurones. En utilisant ces structures, on peut modéliser de manière naturelle les relations de *similarité*. La notion de similarité est cruciale pour la compréhension de nombreux phénomènes cognitifs. J'appellerai *forme conceptuelle* la façon dont je propose de représenter l'information, puisque je pense que l'utilisation de ce type de représentation permet d'obtenir la meilleure description des aspects essentiels de la formation des concepts.

Une fois encore, on ne doit pas considérer que les représentations conceptuelles sont des représentations concurrentes des représentations symboliques ou associationnistes connexionnistes. On doit plutôt les envisager comme trois *niveaux* de représentation de la cognition qui possèdent des échelles de résolution différentes.

Je vais esquisser une théorie des *espaces conceptuels*, qui fournira un cadre particulier à l'intérieur duquel représenter l'information à un niveau conceptuel. Un espace conceptuel est formé à partir de représentations géométriques fondées sur un certain nombre de *dimensions qualitatives*. La théorie mettra l'accent sur l'aspect constructif des sciences cognitives. Je pense toutefois qu'elle peut aussi expliquer un certain nombre d'aspects de nos connaissances sur les représentations dans divers systèmes biologiques.

2. DE L'ORIGINE DES DIMENSIONS QUALITATIVES

A titre d'exemples introductifs de dimensions qualitatives, on peut mentionner la température, le poids, la luminosité, la hauteur sonore et les trois dimensions spatiales ordinaires que sont la hauteur, la largeur et la profondeur. J'ai choisi ces exemples parce qu'ils sont étroitement liés à ce que nos récepteurs sensoriels produisent (Schiffman 1982). Les dimensions spatiales de hauteur, largeur et profondeur ainsi que la luminosité sont perçues par le système visuel, la hauteur par le système auditif, la température par les capteurs thermiques, et le poids enfin, par les capteurs kinesthésiques. Toutefois, certaines dimensions qualitatives possèdent aussi un caractère abstrait non-sensoriel.

Il semble qu'il y ait différents types de dimensions, par conséquent la question se pose : d'où viennent les dimensions ? Je ne pense pas que cette question puisse recevoir une réponse unique. Dans cette section, j'essaierai de repérer les origines de différents types de dimensions qualitatives.

Premièrement, certaines dimensions qualitatives semblent être innées ou développées très tôt. Elles sont dans une certaine mesure câblées dans notre système nerveux, comme c'est le cas par exemple des dimensions sensorielles. Ceci est aussi probablement vrai de nos représentations de l'espace ordinaire. Puisqu'il est évident que des domaines de ce genre sont extrêmement importants pour des activités élémentaires comme trouver de la nourriture, éviter le danger et se déplacer dans l'environnement, l'hypothèse de l'innéité de ces dimensions possède une justification évolutionnaire. Les êtres humains et autres animaux qui n'auraient pas été dotés d'une représentation suffisamment adéquate de la structure spatiale du monde extérieur auraient été désavantagés par la sélection naturelle.

Le cerveau des êtres humains et des animaux contient des aires topographiques qui font correspondre différents types de modalités sensorielles à des aires spatiales. Les principes qui structurent ces correspondances sont en grande partie innés, même si leur mise au

point est établie et affinée pendant le développement de l'humain ou de l'animal. Les mêmes principes semblent gouverner la plupart du règne animal. Selon Gallistel (1990, p. 105) :

[...] la croyance intuitive selon laquelle les cartes cognitives des animaux "inférieurs" sont plus faibles que les nôtres est sans fondement. Elles sont peut-être appauvries relativement aux nôtres (elles sont peut-être moins détaillées), mais elles ne sont pas plus faibles du point de vue de leurs caractéristiques formelles. Il y a des preuves expérimentales qui montrent que même les cartes cognitives que possèdent les insectes sont des cartes métriques.

Quine note que quelque chose comme des dimensions qualitatives innées sont nécessaires pour rendre l'*apprentissage* possible :

Sans une telle capacité à situer les qualités dans l'espace, nous ne pourrions jamais acquérir une habitude, tous les stimuli seraient également semblables et également différents. On peut explorer cette capacité en laboratoire, tant chez l'homme que chez l'animal, au moyen d'expériences sur le conditionnement et l'extinction. Dans la mesure où elles sont nécessaires à tout apprentissage, ces localisations distinctives de qualités dans l'espace ne peuvent elles-mêmes être toutes apprises, certaines d'entre elles doivent être innées. (Quine 1969 : 123)

Néanmoins, une fois que le processus a commencé, l'apprentissage peut contribuer à développer de nouvelles dimensions.¹ Des études sur le développement cognitif de l'enfant fournissent un exemple de ce genre. Des enfants de deux ans peuvent représenter des objets dans leur globalité, mais ils sont incapables de raisonner sur les dimensions de l'objet.

L'apprentissage de nouveaux concepts est par conséquent souvent lié au fait d'*étendre* son espace conceptuel à de nouvelles dimensions qualitatives. Considérons par exemple la dimension (phénoménale) du *volume*. Les expériences sur la "conservation" menées par Piaget et ses successeurs indiquent que les jeunes enfants sont dépourvus d'une dimension mentale séparée de volume. Ils confondent le volume d'un liquide avec la *hauteur* du liquide dans son récipient. C'est seulement à l'âge de cinq ans environ qu'ils apprennent à représenter ces deux dimensions séparément. De même, les enfants de trois et quatre ans confondent *haut* et *grand*, *gros* et *brillant*. Selon Smith (1989, p. 146-47) :

l'élaboration d'un système de dimensions perceptives, d'un système de *types* de similarités, est peut-être l'une des réalisations intellectuelles majeures de la petite enfance. [...] La notion

¹On doit noter qu'il est impossible de faire une distinction tranchée entre les dimensions qualitatives innées et les dimensions qualitatives apprises, puisque beaucoup de dimensions sensorielles sont structurellement préparées dans les tissus nerveux à la naissance, mais exigent une exposition aux expériences sensorielles pour que la structure géométrique exacte de ces correspondances soit déployée.

développementale élémentaire est celle de différenciation, qui s'exerce sur des classes syncrétiques globales de ressemblances et de grandeurs perceptives pour aboutir à des types dimensionnellement spécifiques d'identité et de grandeur.

D'autres dimensions encore peuvent dépendre de la *culture*.² Considérons la dimension du "temps" par exemple. Dans certaines cultures, le temps est conçu comme étant *circulaire* - le monde ne cesse de revenir au même point dans le temps, et les mêmes événements se produisent constamment à nouveau. Dans d'autres cultures, parler du temps comme d'une dimension n'a guère de sens. Une dimension temporelle sophistiquée, possédant une structure métrique complète, est nécessaire pour les formes avancées de planification et de coordination avec les autres individus, mais elle n'est pas nécessaire pour les activités les plus élémentaires d'un organisme. En fait la conception occidentale courante du temps est un phénomène comparativement récent (cf. Toulmin et Goodfield (1965)).

Les exemples présentés ici indiquent qu'un grand nombre de dimensions qualitatives des espaces conceptuels humains ne sont pas directement produites à partir des entrées sensorielles. Ce point est encore plus clair lorsqu'on utilise des concepts qui sont fondés sur la *fonction* des artefacts ou sur les *rôles sociaux* des personnes dans une société. Même si nous ne savons que peu de choses sur la structure géométrique de ces dimensions, il est tout à fait évident qu'une structure non triviale de ce genre existe. Marr et Vaina (1982) et Vaina (1983) ont donné des arguments allant dans ce sens. Ils fournissent une analyse de la représentation fonctionnelle dans laquelle la fonction d'un objet est déterminée en termes des actions qu'il permet.

La culture, sous la forme d'une interaction entre les personnes, peut en elle-même engendrer des contraintes sur les espaces conceptuels. Freyd (1983) par exemple propose de manière très intéressante que les espaces conceptuels ont pu évoluer comme une forme représentationnelle dans une communauté parce que les personnes doivent *partager* des connaissances :

On a proposé un certain nombre d'approches différentes pour analyser les structures en domaines sémantiques, mais le but commun de ces approches est de mettre au jour des contraintes sur la représentation des connaissances. Je soutiens que les structures que les analyses sémantiques permettent de découvrir pourraient provenir de contraintes de partageabilité sur la représentation des connaissances.

²Je ne soutiens pas que ma typologie des origines des dimensions qualitatives soit exclusive, puisqu'en un sens toutes les dimensions qui dépendent de la culture sont aussi apprises.

[...]

Par conséquent, si l'on peut montrer qu'un ensemble de termes se comporte comme s'ils étaient représentés dans un espace tridimensionnel, on fait souvent l'inférence que cette réalité spatiale (ou une formulation équivalente) possède une réalité psychologique et qu'il y a une nécessité innée pour celle-ci. Mais il se pourrait que les propriétés structurales du domaine de connaissance soient présentes parce que de telles propriétés structurales constituent la façon la plus efficace de partager les concepts. C'est-à-dire qu'on ne peut être sûr que ces régularités nous disent quelque chose sur la façon dont le cerveau peut représenter les choses, ou même "préfère" le faire, s'il n'avait pas à partager des concepts avec d'autres cerveaux". (1983, pp. 193-194)

Freyd suggère ici une explication *économique* de la raison pour laquelle nous avons des espaces conceptuels : ils facilitent le partage des connaissances.

Enfin, certaines dimensions qualitatives sont introduites par la *science*. En témoigne par exemple la distinction introduite par Newton entre *poids* et *masse*, qui est d'une importance cruciale pour le développement de sa mécanique céleste, mais qui n'a guère de correspondant dans la perception humaine. Pour autant que nous ayons des représentations mentales de la masse des objets par opposition à leur poids, elles ne nous sont pas données par les sens, mais doivent être apprises en adoptant l'espace conceptuel de la mécanique newtonienne dans nos représentations.

Les changements les plus drastiques en science se produisent lorsque l'espace conceptuel sous-jacent est modifié. Je pense que la plupart des "changements de paradigme" discutés par Kuhn (1972) peuvent être compris comme des *changements d'espaces conceptuels*. Je ne vois aucune différence de principe entre ce type de changement et le changement impliqué dans le développement de l'espace conceptuel d'un enfant. L'introduction de la distinction entre "hauteur" et "volume" constitue le même genre de phénomène que l'introduction par Newton de la distinction entre "poids" et "masse". Cette distinction est de nos jours omniprésente en physique, même si elle n'a guère de base sensorielle.

On peut maintenant définir un *espace conceptuel* comme une collection d'une ou de plusieurs dimensions qualitatives. On ne doit cependant pas considérer les dimensions de l'espace conceptuel comme des entités totalement indépendantes. Elles sont en effet *corrélées* de diverses manières, puisque les propriétés des objets modélisés dans cet espace covarient. Par exemple, la dimension de la maturité et celle de la couleur covarient dans l'espace des fruits.

3. LA FORMATION DE CONCEPTS DECRITE A L'AIDE DES ESPACES CONCEPTUELS

En des termes plus abstraits, un espace conceptuel S consiste en une classe D_1, \dots, D_n de dimensions qualitatives. Un point dans S est représenté par un vecteur $v = \langle d_1, \dots, d_n \rangle$ avec un indice pour chaque dimension. Chacune des dimensions est dotée d'une certaine structure topologique ou métrique. Le but de cette section est de montrer comment on peut utiliser les espaces conceptuels pour modéliser les *concepts*.

En première approximation, on peut décrire un concept comme une *région* dans l'espace conceptuel S , où l'on doit comprendre 'région' comme une notion spatiale déterminée par la topologie et la métrique de S . Par exemple, le point dans la dimension du temps qui représente 'maintenant' divise cette dimension, et ainsi l'espace des vecteurs, en deux régions correspondant aux concepts 'passé' et 'futur'. Mais cette proposition souffre d'un manque de précision en ce qui concerne la notion de 'région'. Une idée plus précise et plus puissante repose sur le critère suivant, où les caractéristiques topologiques des dimensions qualitatives sont utilisées pour introduire une structure spatiale sur les concepts :

Critère CN : un *concept naturel* est une région convexe d'un espace conceptuel.

Une région *convexe* est caractérisée par le critère suivant : pour toute paire de points v_1 et v_2 dans la région, tous les points situés entre v_1 et v_2 sont aussi dans la région. Ce qui motive l'introduction de ce critère est le fait que si certains objets qui sont situés à v_1 et v_2 relativement à une certaine dimension qualitative (ou plusieurs d'entre elles) sont tous deux des exemples du concept C , alors tout objet qui est situé entre v_1 et v_2 sur les dimensions qualitatives sera aussi un exemple de C . Je montrerai plus loin que ce critère est psychologiquement réaliste. Le critère CN suppose que la notion du *fait d'être-situé-entre* a un sens pour les dimensions qualitatives pertinentes. Ceci est cependant une hypothèse plutôt faible, qui n'exige que très peu de structure topologique sous-jacente.

La plupart des concepts exprimés par des mots simples dans les langues naturelles sont des concepts naturels dans le sens que je viens de spécifier. Par exemple, je fais la conjecture que tous les *termes de couleur* présents dans les langues naturelles expriment des concepts naturels relativement à la représentation psychologique des trois dimensions des couleurs, ceci veut dire nuance, intensité et luminosité. En d'autres termes, la conjecture prédit que si un certain objet o_1 est décrit par le terme de couleur C dans une langue donnée, et si un objet o_2 est dit aussi avoir la couleur C , alors tout objet o_3 dont la couleur est située entre celle de o_1 et celle de o_2 sera aussi décrit par le terme de couleur C . Il est bien connu que des langues différentes découpent le cercle des couleurs de manière différente, mais il semble que tous ces découpages soient faits en termes d'ensembles convexes. Berlin et Kay (1969) fournissent des éléments

qui étayent fortement cette conjecture, bien qu'ils ne traitent pas des termes de couleur en général, mais se concentrent sur les termes de couleurs élémentaires.

L'identification phonétique des *voyelles* dans diverses langues fournit une autre illustration de la façon dont la convexité des régions détermine les concepts et la catégorisation. D'après la théorie phonétique, une voyelle est déterminée par les relations entre la fréquence élémentaire du son et ses formants (des hautes fréquences qui sont présentes en même temps que la fréquence élémentaire). En général, les deux premiers formants F_1 et F_2 suffisent à identifier une voyelle. Ceci veut dire qu'on peut utiliser les coordonnées de l'espace bidimensionnel délimités par F_1 et F_2 (relativement à une hauteur élémentaire fixée F_0) pour décrire une voyelle de manière assez exacte. Fairbanks et Grubb (1961) ont étudié la façon dont les sujets produisent et reconnaissent les voyelles en Américain parlé. La figure 1 résume leurs résultats.

Il est important de noter que dans cet exemple, le fait d'identifier F_1 et F_2 comme les dimensions pertinentes de la formation des voyelles est une *découverte* phonétique. Nous possédions déjà le concept de voyelle avant cette découverte, mais l'analyse spatiale nous permet de comprendre plusieurs caractéristiques de la classification des voyelles dans différentes langues.

Le critère CN fournit une théorie des concepts qui est indépendante à la fois des mondes possibles et des individus. Il satisfait en outre le desideratum de Stalnaker selon lequel un concept "... ne doit pas seulement être une règle de groupement des individus, mais une caractéristique des individus en vertu de laquelle ils peuvent être groupés" (Stalnaker 1981, p. 347). Je voudrais toutefois souligner que je considère ce critère seulement comme une condition *nécessaire*, mais peut-être pas suffisante, d'un concept naturel. Ce critère délimite une classe de concepts qui sont utiles à des fins cognitives, mais il se pourrait qu'il ne soit pas suffisamment restrictif.

³ En revanche, la référence à un terme artificiel de couleur comme 'grue' (Goodman 1984) ne sera pas une région convexe dans l'espace conceptuel ordinaire, et donc n'est pas un concept naturel selon le critère CN. Pour une analyse approfondie de cet exemple, cf. Gärdenfors (1990).

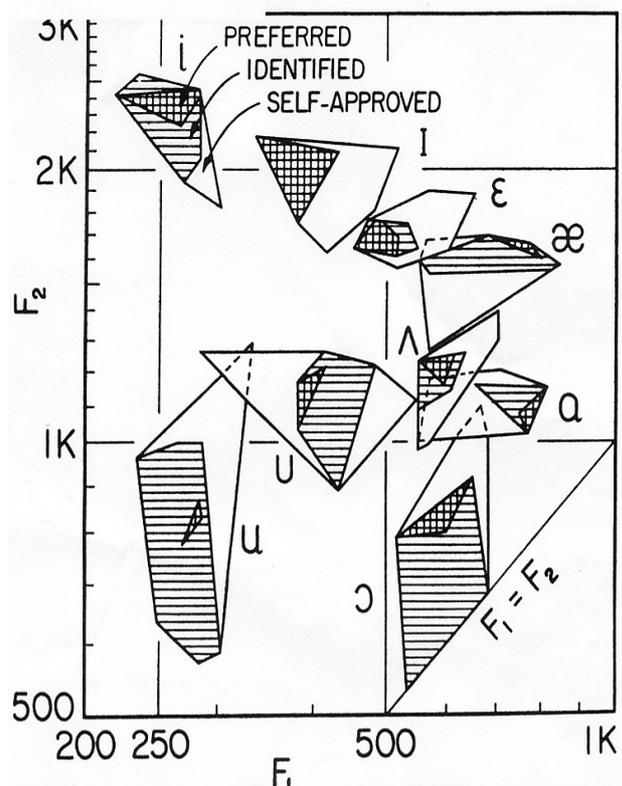


Figure 1.

L'échelle des abscisses et des ordonnées est le logarithme des fréquences de F_1 et F_2 (la fréquence F_0 de base des voyelles était 130 Hz). Comme on peut le voir à partir du diagramme, les exemples préférés, identifiés et auto-approuvés des différentes voyelles forment des sous-régions convexes de l'espace déterminé par F_1 et F_2 avec l'échelle logarithmique donnée.⁴ Comme c'est le cas des termes de couleur, différentes langues découpent l'espace phonétique de manière différente (le nombre de voyelles identifiées dans différentes langues varie considérablement), mais je conjecture à nouveau que chaque voyelle à l'intérieur d'une langue correspondra à une région convexe de l'espace des formants.

⁴Une voyelle *auto-approuvée* est une voyelle qui a été produite par le locuteur et plus tard approuvée comme un exemple du type en question. Un échantillon *identifié* d'une voyelle est un échantillon qui a été correctement identifié par 75% des observateurs. Les échantillons *préférés* d'une voyelle sont ceux qui sont "les plus représentatifs parmi les échantillons les plus facilement identifiés" (Fairbanks et Grubb 1961, p. 210).

4. LES DIMENSIONS QUALITATIVES

La fonction première des dimensions qualitatives est de représenter diverses "qualités" des objets.⁵ Elles correspondent aux différentes façons dont on juge que des stimuli sont similaires ou différents. Dans la plupart des cas, les jugements de similarité et de différence engendrent une relation d'ordre sur les stimuli (Clark 1993, p. 114). Par exemple, on peut juger les sons par leur hauteur, ce qui engendrera un ordre sur les perceptions. Les dimensions forment le "cadre" utilisé pour assigner des *propriétés* aux objets et pour spécifier des *relations* entre eux. Les coordonnées d'un point à l'intérieur d'un espace conceptuel représentent des instanciations particulières de chaque dimension, par exemple une température particulière, un poids particulier, etc.

Je considère que les dimensions qualitatives sont indépendantes des représentations symboliques en ce sens que les humains et les autres animaux représentent les qualités des objets, par exemple lors de la planification d'une action, sans qu'il soit nécessaire de présupposer un langage interne ou un autre système symbolique dans lequel ces qualités sont exprimées. En d'autres termes, les dimensions sont les éléments primitifs des représentations au niveau conceptuel.

Si l'on se concentre sur le but explicatif des sciences cognitives, il faut considérer les dimensions qualitatives comme des entités théoriques qui servent de facteur modélisateur dans la description des capacités cognitives d'un organisme. Lorsqu'on construit des systèmes artificiels, les dimensions jouent le rôle d'un cadre pour les représentations utilisées par le système.

Il faut comprendre littéralement la notion de dimension. On fait l'hypothèse que chacune des dimensions qualitatives est dotée de certaines structures *géométriques* (dans certains cas, il s'agit de structures *topologiques* et de structures d'*ordre*). La dimension de "poids" fournit un premier exemple pour illustrer une telle structure (cf. figure 2). Il s'agit d'une structure unidimensionnelle possédant un point zéro, elle est donc isomorphe à la demi-droite des nombres non-négatifs. On pose communément en science une contrainte sur cette dimension, selon laquelle il n'y a pas de poids négatif.⁶

⁵Dans la philosophie traditionnelle, on fait souvent, à la suite de Locke, la distinction entre qualités premières et qualités secondes. Cette distinction correspond en gros à la distinction entre la dimension "scientifique" et la dimension "phénoménale" qui sera introduite dans la section suivante.

⁶Il est toutefois intéressant de noter (cf. Kuhn 1972) que pendant la période de la chimie du phlogistique, les scientifiques prenaient en compte des poids négatifs afin d'éviter certaines anomalies qui affectaient la théorie.



Figure 2. La dimension des poids

Dans un texte précédent sur les espaces conceptuels, j'ai utilisé l'exemple de l'espace perceptif des couleurs afin d'illustrer un ensemble plus structuré de dimensions qualitatives (Gärdenfors 2000). On peut toutefois trouver d'autres structures spatiales du même genre pour d'autres qualités sensorielles. Considérons par exemple la dimension qualitative de *hauteur* de ton qui est essentiellement une structure continue unidimensionnelle allant des sons les plus graves aux sons les plus aigus. Cette représentation est directement liée à la neurophysiologie de la perception de la hauteur.

En plus de la dimension qui se rapporte à la fréquence de base des sons, il est aussi possible d'identifier d'autres structures à l'intérieur la représentation mentale des sons. Les sons naturels ne sont pas de simples sons sinusoïdaux possédant une fréquence unique, mais ils sont constitués par un certain nombre d'harmoniques de rang plus élevé. Le timbre d'un son, qui est une dimension phénoménale, est déterminé par la force relative des harmoniques de rang plus élevé de la fréquence fondamentale d'un son. Le "cas de la fondamentale absente" est un phénomène perceptif intéressant. Si au moyen d'une méthode artificielle, on élimine la fréquence fondamentale d'un son physique complexe, on continue néanmoins à percevoir la hauteur phénoménale du son comme correspondant à la fondamentale qui a été éliminée.⁷ Apparemment, la fréquence fondamentale n'est pas indispensable à la perception de la hauteur, mais la hauteur perçue est déterminée par une combinaison des harmoniques de rang supérieur.

Ainsi, les harmoniques d'un son sont essentiels à la façon dont il est perçu. Ceci implique que les sons qui partagent un certain nombre d'harmoniques sont perçus comme similaires. Le son qui partage le plus d'harmoniques avec un son donné est son octave, viennent ensuite la quinte, la quarte, et ainsi de suite. Cette structure "géométrique" additionnelle sur la dimension de la hauteur, qui peut être dérivée de la structure ondulatoire des sons, fonde l'explication de la perception des *intervalles* musicaux.⁸

Pour donner un autre exemple de représentation d'un espace sensoriel, je mentionnerai seulement que la perception humaine du *goût* semble être produite par quatre types distincts de récepteurs : salé, acide, sucré et amer. On peut ainsi décrire l'espace qualitatif représentant les goûts comme un espace en quatre dimensions.

⁷Voir par exemple Gabrielsson (1981), pp. 20-21.

⁸Pour une plus ample discussion de la structure de l'espace musical, cf. Gärdenfors (1988), sections 7-9.

Henning (1916) a proposé un modèle de ce genre. Il a suggéré qu'on pouvait décrire l'espace gustatif phénoménal comme un tétraèdre (cf. figure 3). En fait Henning avait émis l'hypothèse qu'on pouvait décrire n'importe quel goût comme une combinaison de seulement trois primitives. Ceci veut dire que n'importe quel goût peut être représenté comme un point sur un des *plans* du tétraèdre, de sorte qu'aucun des goûts ne correspond à un point situé à l'intérieur du tétraèdre.

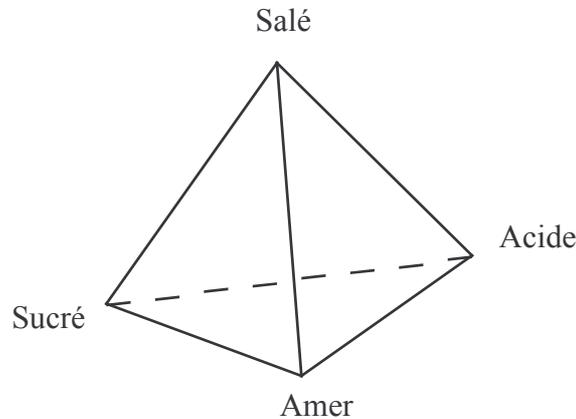


Figure 3. Le tétraèdre du goût de Henning

D'autres modèles toutefois proposent plus de quatre goûts fondamentaux.⁹ Le meilleur modèle de l'espace gustatif phénoménal reste encore à inventer. Ce projet mettra en jeu des techniques de mesure psychophysique sophistiquées. Il suffit de dire qu'il est tout à fait clair que l'espace gustatif possède une structure géométrique non triviale. On peut par exemple affirmer que le goût d'une noix est *plus proche* du goût d'une noisette que du goût du pop-corn, de même qu'on peut dire que la couleur orange est plus proche du jaune que du bleu.

On doit noter que certaines "dimensions" qualitatives ont seulement une structure *discrète*, c'est-à-dire qu'elles divisent simplement les objets en classes disjointes. Deux exemples de ce cas sont les classifications des espèces biologiques et des relations de parenté dans une société humaine. La figure 4 présente un exemple d'arbre phylogénétique tel qu'on les trouve en biologie. Les nœuds représentent des espèces différentes dans l'évolution d'une famille d'organisme par exemple, et les nœuds situés plus haut sur l'arbre représentent des espèces évolutionnairement plus anciennes (éteintes).

⁹Cf. Schiffman (1982), chapitre 9 pour un exposé de certaines de ces théories.

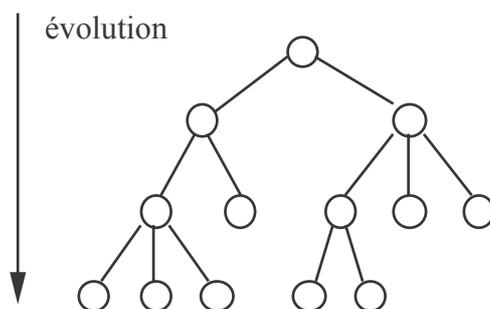


Figure 4. Arbre phylogénétique

On peut mesurer la distance entre deux nœuds par la longueur du chemin qui les relie entre eux. Ceci veut dire que même dans le cas de dimensions discrètes, on peut distinguer une structure géométrique rudimentaire. Par exemple, dans la classification phylogénétique des animaux, il y a un sens à dire que les oiseaux et les reptiles sont *plus étroitement apparentés* que ne le sont les reptiles et les crocodiles.

L'importance de la notion de *similarité* est sérieusement minimisée par l'approche symbolique des représentations. Je suggère que les jugements de similarité soient centraux pour un grand nombre de processus cognitifs. Les jugements de similarité révèlent les dimensions de nos perceptions et leur structure. Pour de nombreux types de dimensions, il sera possible de parler de *distances*. L'hypothèse générale est que plus la distance entre la représentation de deux objets est faible, plus les représentations sont similaires. De cette manière, on peut définir la similarité de deux objets par la distance entre les points qui les représentent dans l'espace. Ainsi, les espaces conceptuels fournissent une manière naturelle de représenter les similarités. De façon générale, le rôle épistémologique des espaces conceptuels est de servir d'outils pour classer diverses *relations* entre les perceptions.

5. LES INTERPRÉTATIONS PHÉNOMÉNALES ET SCIENTIFIQUES DES DIMENSIONS

Il est important d'introduire une distinction entre une interprétation *phénoménale* (ou *psychologique*) et une interprétation *scientifique*. L'interprétation phénoménale porte sur les structures cognitives (perceptions, souvenirs, etc.) des êtres humains et des autres organismes. L'interprétation scientifique, de son côté, traite les dimensions comme une partie d'une théorie scientifique.

Cette distinction est pertinente pour les deux buts des sciences cognitives que j'ai présentés en commençant. Lorsqu'on considère les dimensions comme des entités cognitives, c'est-à-dire lorsque le but est d'expliquer des processus cognitifs naturels, leur structure géométrique ne doit pas être déterminée par des théories

scientifiques qui tentent de donner une description "réaliste" du monde, mais par des mesures *psychophysiques* qui déterminent la structure de la façon dont nos perceptions sont représentées. En outre, lorsqu'il s'agit de fournir une sémantique d'une langue naturelle, on s'intéresse aux interprétations phénoménales des dimensions qualitatives.

En revanche, lorsque ce qui est en question est la *construction* d'un système artificiel, la fonction des capteurs, des effecteurs, et des divers dispositifs de contrôle est en général décrite en termes de dimensions modélisées scientifiquement. Par exemple, les variables d'entrée d'un robot peuvent consister en un petit nombre de grandeurs mesurées physiquement, comme la luminosité, le temps mis par un écho radar pour revenir à sa source ou la pression d'une pince mécanique. À l'aide des buts programmés d'un robot, on peut alors transformer ces variables en un certain nombre de grandeurs de sorties physiques, comme par exemple le voltage des moteurs contrôlant la roue droite et la roue gauche du robot.

Prenons l'exemple des couleurs pour illustrer la distinction les deux interprétations. La distinction que je viens d'introduire est étayée par Gallistel (1990, p. 518-519) qui écrit :

Les faits sur la vision des couleurs suggèrent à quel point le système nerveux peut être porté à représenter des stimuli comme des points dans des espaces descriptifs de dimensionalité modeste. Il en va ainsi même dans le cas de la composition spectrale, qui ne se prête pas à une telle représentation. Le manque de correspondance qui en résulte entre la représentation psychologique de la composition spectrale et la composition spectrale elle-même est une source de confusion et d'erreurs dans la discussion scientifique des couleurs. Les scientifiques persistent à faire référence aux caractéristiques physiques du stimulus et à la sensibilité caractéristique des transducteurs (les cônes) comme si les termes psychologiques de couleur comme *rouge*, *vert* et *bleu* avaient une traduction directe dans la réalité physique, alors qu'en fait ce n'est pas le cas.

On doit prendre sérieusement l'avertissement de Gallistel contre la confusion entre ces deux types de représentation et les erreurs auxquelles elle donne lieu. Il est en effet très facile de confondre ce que dit la science sur les caractéristiques de la réalité et notre perception de cette réalité. Dans cet article, nous nous intéresserons à la représentation phénoménale.

Pour donner un autre exemple, l'espace conceptuel de la mécanique des particules de Newton est bien entendu fondé sur des dimensions qualitatives scientifiques (théoriques), et non sur des dimensions phénoménales (psychologiques). Les dimensions qualitatives de cette théorie sont l'espace ordinaire (l'espace euclidien tridimensionnel), le temps (isomorphe aux nombres réels), la masse (isomorphe aux nombres réels non négatifs), et la force (espace euclidien tridimensionnel). Une fois qu'on a assigné une valeur à une particule pour ces huit dimensions, elle est complètement décrite du

point de vue de la mécanique newtonienne. Dans cette théorie, un objet est donc représenté comme un point dans un espace à huit dimensions.

6. RELATION A LA THEORIE DES PROTOTYPES

L'idée qui consiste à décrire les concepts comme des régions convexes des espaces conceptuels s'accorde très bien avec la *théorie des prototypes* de la catégorisation développée par Rosch et ses collaborateurs (Rosch 1975, 1978, Mervis et Rosch 1981, Lakoff 1987). L'idée principale de la théorie des prototypes est qu'entre les membres d'une catégorie d'objets, comme ceux quiinstancient un concept, on juge que certains membres sont plus représentatifs de la catégorie que d'autres. Par exemple, on considère que les rouges-gorges sont plus représentatifs de la catégorie 'oiseau' que ne le sont les corbeaux, les pingouins ou les émeus ; et les chaises de bureau sont des instanciations plus typiques de la catégorie 'chaise' que ne le sont les rocking chairs, les chaises-longues ou les poufs. Les membres les plus représentatifs d'une catégorie sont appelés membres *prototypiques*. Il est bien connu que certains concepts, comme 'rouge' ou 'chauve' n'ont pas de limites tranchées, et dans leur cas, il n'est peut-être pas surprenant de trouver des effets prototypiques. On a cependant mis en évidence des effets de ce genre pour la plupart des concepts, y compris ceux qui possèdent des limites comparativement claires, comme 'oiseau' et 'chaise'.

Dans le cadre des analyses philosophiques traditionnelles des concepts, fondées sur les fonctions de vérité ou les mondes possibles, il est très difficile d'expliquer de tels effets prototypiques (cf. Gärdenfors 1991). Ou bien un objet est un élément de la classe assignée à un concept (relativement à un monde possible donné), ou bien il ne l'est pas, et tous les éléments de la classe ont un statut égal en tant que membres de la catégorie. Les recherches de Rosch ont visé à faire apparaître l'existence d'asymétries entre les membres d'une catégorie et des structures asymétriques à l'intérieur des catégories. Puisque la définition traditionnelle du concept ne prédit ni n'explique de telles asymétries, on doit trouver le moyen d'en rendre compte en adoptant une notion de concept différente.

Si l'on décrit les concepts comme des régions convexes d'un espace conceptuel, on doit en revanche s'attendre à trouver des effets prototypiques. Dans une région convexe, on peut décrire des positions comme étant plus ou moins *centrales*. Par exemple, si des concepts de couleur sont identifiés à des sous-ensembles convexes de l'espace des couleurs, les points centraux de ces régions seraient les exemples les plus prototypiques de la couleur. Dans une série d'expériences, Rosch a pu démontrer la réalité psychologique de telles couleurs 'focales'. La catégorisation des voyelles présentée dans la section précédente fournit une autre illustration de ce point. Dans ce cas aussi, la structure des différents types de réponses données par les sujets manifeste des effets prototypiques clairs.

Dans le cas de catégories plus complexes comme 'oiseau', il est peut-être plus difficile de décrire l'espace conceptuel sous-jacent. Cependant si l'on adopte quelque chose comme l'analyse des formes proposée par Marr et Nishihara (1978), on peut commencer à voir comment un tel espace pourrait apparaître.¹⁰ Le schéma qu'ils proposent pour décrire les formes biologiques utilise des hiérarchies de formes primitives cylindriques. Chaque cylindre est décrit par deux coordonnées (longueur et largeur). Les cylindres sont combinés en déterminant l'angle entre le cylindre dominant et celui qui est ajouté (deux coordonnées polaires) et la position du cylindre ajouté en relation au cylindre dominant (deux coordonnées). Les détails de la représentation ne sont pas importants dans le présent contexte, mais il faut noter qu'à chaque niveau de la hiérarchie, un objet est décrit par un nombre comparativement petit de coordonnées basées sur les longueurs et les angles. Ainsi, l'objet peut être identifié comme un vecteur hiérarchiquement structuré dans un espace conceptuel (d'ordre supérieur). La figure 5 illustre la structure hiérarchique de leur représentation.

Il faut noter que même si l'on juge que des membres différents d'une certaine catégorie sont plus ou moins prototypiques, il ne s'ensuit pas qu'un objet existant doit représenter 'le prototype'. On peut aisément expliquer ce point si l'on adopte la thèse qu'un concept est une région convexe d'un espace conceptuel, puisque l'élément central de la région (s'il est unique) est un individu possible dans le sens discuté plus haut (si toutes ses dimensions sont spécifiées), mais il n'est pas nécessaire qu'il fasse partie des éléments existants de la catégorie. Il n'est pas nécessaire qu'un tel point prototypique dans la région soit complètement décrit comme un individu, mais il est normalement représenté comme un vecteur partiel, où seules les valeurs des dimensions qui sont pertinentes pour le concept ont été déterminées. Par exemple, la forme générale de l'oiseau prototypique serait incluse dans le vecteur, mais on peut présumer que sa couleur ou son âge ne le seraient pas.

Il est aussi possible de fournir des arguments en faveur de la réciproque, et de montrer que si l'on adopte la théorie des prototypes, alors on doit s'attendre à ce que les concepts soient représentés comme des régions convexes. Faisons l'hypothèse que certaines dimensions qualitatives d'un espace conceptuel soient données, par exemple les dimensions de l'espace des couleurs, et supposons que nous voulions le diviser en un certain nombre de catégories, par exemple les catégories de couleur. Si nous partons d'un ensemble de prototypes p_1, \dots, p_n des catégories, par exemple les couleurs focales, alors ceux-ci devraient être le point central dans les catégories qu'ils représentent. Une façon d'utiliser cette information est de faire l'hypothèse que pour tous les points p dans l'espace, on peut mesurer la *distance* entre p et chacun des p_i . Si maintenant on

¹⁰Cette analyse est étendue dans Marr (1982), ch. 5. Biederman (1987) présente un modèle en rapport ainsi que certains fondements psychologiques.

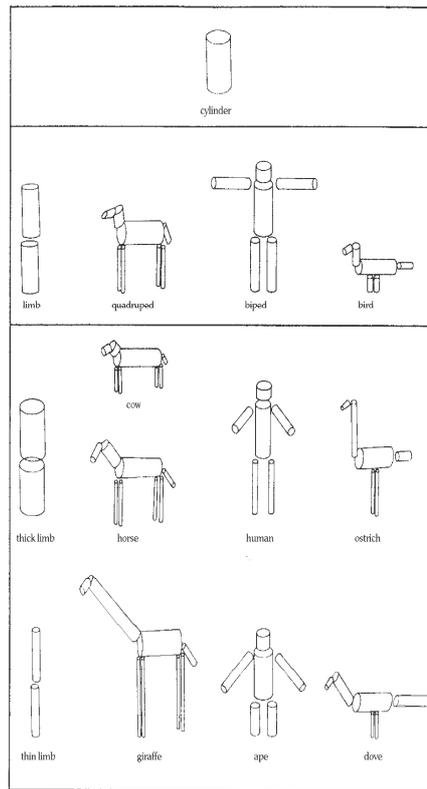


Figure 5.

stipule que p appartient à la même catégorie en tant que le prototype p_i le plus proche, on peut montrer que cette règle engendrera une partition de l'espace qui *consiste en régions convexes* (on définit ici la convexité en termes d'une mesure de distance supposée). C'est ce qu'on appelle la partition de Voronoi dont la figure 6 fournit un exemple bidimensionnel.

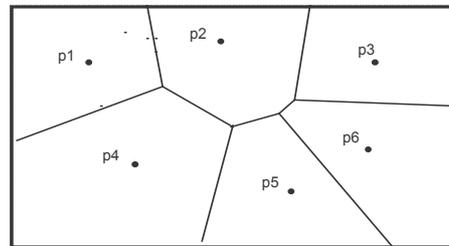


Figure 6.

Ainsi, si l'on fait l'hypothèse qu'une métrique est définie sur le sous-espace qui est soumis à la catégorisation, un ensemble de prototypes engendrera par cette méthode une partition unique du sous-espace en régions convexes. Il y a donc un lien intime entre la théorie des prototypes et l'analyse proposée dans cet article, où les concepts sont décrits comme des régions convexes dans un espace conceptuel.

Le travail de Petitot (1989) offre un exemple concret de cette technique. Il applique la catégorisation de Voronoi pour expliquer certains aspects de la perception catégorique des phonèmes. Il analyse en particulier la relation entre les consonnes occlusives /b/, /d/, /g/, /p/, /t/, /k/. Les relations entre ces consonnes sont exprimées à l'aide de deux dimensions : la première est la dimension voisé-non-voisé, la seconde est constituée par les dimensions labial-dental-velaire qui se rapportent au lieu de l'articulation de la consonne. Ces deux dimensions peuvent être traitées comme continues. La figure 7 montre comment Petitot représente les limites entre les consonnes.

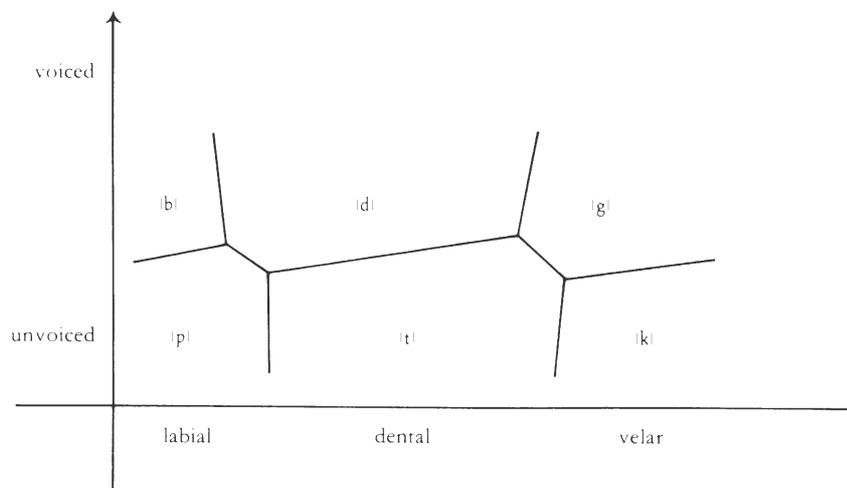


Figure 7. Un modèle Voronoi des limites entre les consonnes occlusives

(Petitot (1989), p. 69)

Petitot (1989, p. 68) explicite l'information que ce modèle contient :

La *géométrie* du système des limites peut fournir des informations précieuses sur les relations hiérarchiques que les consonnes occlusives entretiennent entre elles. Le fait que dans le modèle de Massaro et Oden le domaine de /p/ et celui de /d/ soit *adjacents*, alors que celui de /b/ et celui de /t/ soient *séparés*, indique que le contraste entre /b/ et /t/ est beaucoup plus grand que celui entre /p/ et /d/.

7. CONCLUSION

Le but principal de cet article était de présenter le cœur de la théorie des espaces conceptuels. A ce sujet, il est important de se demander quel *genre* de théorie est la théorie des espaces conceptuels. Est-ce une théorie empirique, une théorie normative, une théorie computationnelle, une théorie psychologique, une théorie neuroscientifique ou une théorie linguistique ?

Comme je l'ai écrit dans la section 1, les sciences cognitives ont deux buts principaux : *expliquer* les phénomènes cognitifs et *construire* des systèmes artificiels qui peuvent résoudre diverses tâches cognitives. La théorie des espaces conceptuels se présente comme un *cadre pour représenter l'information*. Elle doit être considérée comme une théorie qui complète le modèle symbolique et le modèle connexionniste, et jette un pont entre ces formes de représentation.

Le but premier est d'utiliser la théorie des espaces conceptuels pour des tâches constructives. Dans de précédents travaux, j'ai montré comment on pouvait l'utiliser à l'intérieur de modèles computationnels de la *formation de concepts* (Gärdenfors 2000) et de l'*induction* (Gärdenfors 1990, 1993, 2000), et j'ai aussi montré qu'elle est utile pour représenter la *signification* de différents types d'expressions linguistiques à l'intérieur d'une approche computationnelle de la sémantique.

Il n'y a pas de limite tranchée entre les usages explicatifs et les usages constructifs des espaces conceptuels. Lorsque, par exemple, on construit le monde représentationnel d'un robot, il vaut souvent la peine de tirer des leçons de la façon dont la biologie a résolu les problèmes dans le cerveau des êtres humains et des autres animaux. Inversement, la construction d'un système artificiel qui peut résoudre un problème cognitif particulier peut fournir des indications sur la façon dont on devrait entreprendre une investigation empirique des systèmes biologiques. Par conséquent, il y a une interaction en spirale entre les usages explicatifs et les usages constructifs des espaces conceptuels.

Cet article a posé des questions au sujet de la géométrie de la pensée. A l'aide de la notion d'espace conceptuel, j'ai fourni une analyse des concepts. La notion de *concept naturel* est une notion clef, qui est définie en termes de régions convexes des espaces conceptuels - une définition qui met cruciallement en jeu la structure géométrique des différents domaines.

A mon avis, un niveau conceptuel de représentation devrait jouer un rôle central dans les sciences cognitives. Après avoir été dominante pendant de nombreuses années, l'approche symbolique a été remise en question par le connexionnisme (qui est de nos jours étendu à une étude plus large des systèmes dynamiques). Dans bien des cas cependant, le niveau symbolique de représentation fournit des analyses trop grossières et le niveau connexionniste des analyses trop fines. Relativement aux deux buts des sciences cognitives que

j'ai énoncés, je pense que le niveau conceptuel contribuera de manière significative à augmenter nos capacités explicatives quand il s'agit de comprendre les systèmes cognitifs, en particulier ceux qui sont liés à la formation de concept et à la compréhension du langage.

Le principal facteur qui empêche un développement rapide des différentes applications des espaces conceptuels est le manque de connaissances au sujet des dimensions qualitatives pertinentes. C'est presque uniquement dans le cas des dimensions perceptives que les recherches en psychophysique sont parvenues à identifier les structures géométriques et topologiques sous-jacentes (et, dans de rares cas, la métrique psychologique). Par exemple, nous n'avons qu'une compréhension limitée de la façon dont nous percevons et conceptualisons les objets selon leur forme.

La découverte de la structure des dimensions d'un domaine particulier conduit souvent à des recherches fécondes. Par exemple, le développement de l'espace des voyelles que j'ai présenté dans la section 3 a conduit à un grand nombre de résultats nouveaux en phonétique, et à une compréhension approfondie du processus de parole.

Par conséquent, ceux qui veulent contribuer à ce programme devraient commencer par partir à la recherche des espaces conceptuels cachés, puisque ces derniers, pour commencer, déterminent la signification des expressions linguistiques. Même si les résultats peuvent être difficiles à obtenir, il est certain qu'ils auront des répercussions sur d'autres domaines des sciences cognitives.

Bibliographie

- Berlin, B. et Kay, P. (1969), *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley, CA.
- Biederman, I. (1987), "Recognition-by-components: a theory of human image understanding," *Psychological Review* 94, 115-147.
- Clark, A. (1993), *Sensory Qualities*, Clarendon Press, Oxford.
- Fairbanks, G. et Grubb, P. (1961), "A psychophysical investigation of vowel formants," *Journal of Speech and Hearing Research* 4, 203-219.
- Freyd, J. (1983), "Shareability: the social psychology of epistemology," *Cognitive Science* 7, 191-210.
- Gabrielsson, A. (1981), "Music psychology – a survey of problems and current research activities," in *Basic Musical Functions and Musical Ability*, Publications issued by the Royal Swedish Academy of Music, No. 32, 7-80.
- Gallistel, C. R. (1990), *The Organization of Learning*, MIT Press, Cambridge, MA.
- Gärdenfors, P. (1988), "Semantics, conceptual spaces and music," in Rantala, V. Rowell, L. and Tarasti, eds., *Essays on the Philosophy of Music (Acta Philosophica Fennica, vol. 43)*, Helsinki, 9-27.
- Gärdenfors, P. (1990), "Induction, conceptual spaces and AI," *Philosophy of Science* 57, 78-95.

- Gärdenfors, P. (1991), "Frameworks for properties: possible worlds vs. conceptual spaces," *Language, Knowledge and Intentionality (Acta Philosophica Fennica, vol. 49)*, ed. by L. Haaparanta, M. Kusch, and I. Niiniluoto, Helsinki, 383-407
- Gärdenfors, P. (1993), "Induction and the evolution of conceptual spaces," in Moore, E. C., ed., *Charles S. Peirce and the Philosophy of Science*, The University of Alabama Press, Tuscaloosa, 72-88.
- Gärdenfors, P. (2000), *Conceptual Spaces: On the Geometry of Thought*, MIT Press, Cambridge, MA.
- Goodman, N. (1984), *Fact, fictions et prédictions*, Minit, Paris.
- Henning, H. (1916), "Die Qualitätenreihe des Geschmacks," *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 74, 203-219.
- Kuhn, T. (1972), *Le structure des révolutions scientifiques*, Flammarion, Paris.
- Lakoff, G. (1987), *Women, Fire, and Dangerous Things*. The University of Chicago Press, Chicago, IL.
- Marr, D. (1982), *Vision*. Freeman, San Francisco.
- Marr, D. et Nishihara, H. K. (1978), "Representation and recognition of the spatial organization of three-dimensional shapes," *Proceedings of the Royal Society in London, B* 200: 269-294.
- Marr, D. et Vaina, L. (1982), "Representation and recognition of the movements of shapes," *Proceedings of the Royal Society in London, B* 214, 501-524.
- Mervis, C. et Rosch, E. (1981), "Categorization of natural objects," *Annual Review of Psychology*, 32, 89-115.
- Petitot, J. (1989), "Morphodynamics and the categorical perception of phonological units," *Theoretical Linguistics* 15, 25-71.
- Quine, W. V. O. (1969), "Natural kinds," in *Ontological Relativity and Other Essays*, Columbia University Press, New York, NY, 114-138.
- Rosch, E. (1975), "Cognitive representations of semantic categories," *Journal of Experimental Psychology: General* 104, 192-233.
- Rosch, E. (1978), "Prototype classification and logical classification: the two systems," in Scholnik, E., ed., *New Trends in Cognitive Representation: Challenges to Piaget's Theory*, Lawrence Erlbaum Associates, Hillsdale, NJ, 73-86.
- Schiffman, H. R. (1982), *Sensation and Perception, 2nd edition*, John Wiley & Sons, New York.
- Smith, L. B. (1989), "From global similarities to kinds of similarities – the construction of dimensions in development," in Vosniadou, S. and Ortony, A., eds. *Similarity and Analogical Reasoning*, Cambridge University Press, Cambridge.
- Stalnaker, R. (1981), "Antiessentialism," *Midwest Studies of Philosophy* 4, 343-355.
- Toulmin, S. et Goodfield, J. (1965), *The Discovery of Time*, Penguin, Harmondsworth.
- Vaina, L. (1983), "From shapes and movements to objects and actions," *Synthese* 54, 3-36.