

**Réseaux de discrimination en psychologie:
L'exemple de CHREST**

Fernand Gobet

ESRC Centre for Research in Development, Instruction and Training

School of Psychology

University of Nottingham

Nottingham NG7 2RD

United Kingdom

email: frg@psyc.nott.ac.uk

+ 44 (115) 951 5402 (phone)

+ 44 (115) 951 5324 (fax)

Remerciements

Cet article a été rendu possible par un support financier de l'Economic and Social Research Council, du Leverhulme Trust, et de l'Université de Nottingham. Ma gratitude va également aux membres du groupe CHREST qui ont contribué de manière importante à cette recherche par leur expertise et par leur enthousiasme, à deux experts scientifiques anonymes pour leur commentaires détaillés, et à Bernard Plancherel pour sa lecture attentive du manuscrit final.

Résumé

Les techniques de modélisation jouent un rôle important dans l'arsenal méthodologique du psychologue. Dans cet article, nous discutons le formalisme des réseaux de discrimination. Utilisant l'architecture CHREST comme exemple, nous démontrons comment ce formalisme peut produire des explications du comportement dans divers domaines, tels que l'acquisition de la langue et de l'expertise. Nous concluons en discutant les avantages et désavantages offerts par les réseaux de discrimination.

Mots clés

apprentissage, apprentissage de la langue, chunking, expertise, mémoire, modélisation, perception, réseau de discrimination

Abstract

Modelling techniques play an important role in psychological research. This article discusses the formalism of discrimination nets. Using the CHREST architecture as an example, we show how this formalism can account for behaviour in various domains, such as the acquisition of language and expertise. We conclude by discussing the advantages and disadvantages offered by discrimination nets.

Keywords

chunking, discrimination network, expertise, language acquisition, learning, modelling, memory, perception

Réseaux de discrimination en psychologie: L'exemple de CHREST

Depuis leur introduction dans les années cinquante, les techniques de modélisation par ordinateur ont occupé une place importante dans l'arsenal méthodologique du psychologue. Ces techniques offrent des avantages certains, en particulier un formalisme précis qui rend nécessaires l'explicitation et la consistance des théories et possible la production de leurs conséquences. En outre, ces techniques permettent de simuler des comportements complexes, et, par rapport à d'autres techniques formelles, telles que les techniques de modélisation mathématique ou statistique, elles ont l'avantage de n'exiger que peu d'hypothèses sur la nature des données.

Comme le notent Newell et Simon (1972), la modélisation par ordinateur est une approche naturelle pour la description dynamique de changements contingents s'opérant dans le temps, en particulier en psychologie, où le recours à des méthodes mathématiques telles que le calcul différentiel atteint rapidement une complexité intraitable, sans parler de la contrainte de tout exprimer sous forme numérique. Ainsi, la modélisation par ordinateur joue le même rôle pour la psychologie que les équations différentielles pour la physique.

De nombreuses méthodes ont été utilisées pour modéliser le comportement; par exemple: réseaux sémantiques, réseaux de neurones, et systèmes de production. Aujourd'hui, les réseaux de neurones (Rumelhart & McClelland, 1986) et les systèmes de production (Soar, Newell, 1990; ACT-R, Anderson, 1993) constituent les méthodes les plus communes. Parmi les techniques plus récentes, on peut mentionner les techniques de "vie artificielle" (Levy, 1992) et les techniques robotiques (Pfeifer & Scheier, 1999). Dans cet article, nous allons considérer une autre approche, la simulation par réseaux de discrimination. Avec cette technique, l'apprentissage et le développement sont modélisés par la construction d'un réseau de discrimination, c.-à-d. d'un réseau de tests portant sur certains aspects des objets à identifier ou à catégoriser.

Le but de cet article est d'illustrer, par des cas précis, les possibilités offertes par les réseaux de discrimination en général, et par CHREST¹ (Gobet, 1993; Gobet & Simon,

2000), en particulier. Nous essayerons d'identifier quelles sont les contributions propres aux réseaux de discrimination, et quelles sont celles propres à d'autres caractéristiques du modèle. Comme la plupart des travaux sur les réseaux de discrimination ont été publiés en anglais, cet article va également servir d'introduction à ces techniques pour la communauté scientifique francophone.

Cet article est organisé en six sections. Premièrement, nous décrirons les caractéristiques principales des réseaux de discrimination et les raisons de leur emploi en psychologie. Deuxièmement, nous présenterons brièvement la théorie d'EPAM (Elementary Perceiver And Memorizer; Feigenbaum & Simon, 1984), une architecture cognitive basée sur la construction de réseaux de discrimination. Nous décrirons ensuite CHREST, une élaboration d'EPAM. Dans la quatrième section, nous discuterons certaines simulations de CHREST dans le domaine de l'expertise et de son acquisition, et constaterons que CHREST explique de nombreux phénomènes perceptifs et mnésiques ainsi que certains aspects de résolution de problème. Dans la cinquième section, nous montrerons comment CHREST simule certains aspects de l'apprentissage de la langue maternelle, en particulier l'acquisition de catégories syntaxiques et celle du vocabulaire (lexique). Tant pour l'acquisition de la langue maternelle que pour l'acquisition de l'expertise, les modèles apprennent en utilisant un matériel reflétant l'environnement naturel du domaine, et les simulations sont confrontées en détail avec les données empiriques. Finalement, nous résumerons les avantages et désavantages offerts par les techniques de réseaux de discrimination, et spéculerons sur de possibles directions de recherche futures.

Modélisation psychologique avec réseaux de discrimination

Un réseau de discrimination est une structure symbolique qui permet d'assigner un stimulus à sa représentation en mémoire. Il consiste en un réseau séquentiel de tests portant sur les caractéristiques des objets à catégoriser, et comporte une hiérarchie de noeuds, avec une racine à son sommet. Les noeuds, qui sont reliés par des arcs, contiennent soit une description partielle (appelée *image*) des objets externes, soit un *test* qui va être appliqué pour choisir l'arc à prendre durant la discrimination, ou les deux (voir

Figure 1). Le réseau est créé dynamiquement, graduellement, et de manière non-supervisée, à mesure que des objets externes sont présentés. Par conséquent, les réseaux de discrimination sont des systèmes auto-organiseurs, qui se développent en fonction des caractéristiques de l'environnement et de l'état actuel du système. Les noeuds de plus en plus complexes qui sont acquis durant l'apprentissage d'un domaine constituent des unités significatives, qui correspondent à la notion de *chunk* (Chase & Simon, 1973). Comme ces noeuds peuvent être récursivement utilisés pour créer de nouveaux noeuds, les réseaux de discrimination effectuent une compression des données.

 Insérer Figure 1

Bien que ne jouissant pas de la popularité des réseaux de neurones ou des systèmes de production, les réseaux de discrimination ont joué un rôle important en intelligence artificielle et en psychologie. En intelligence artificielle, ils ont été en particulier utilisés dans l'apprentissage par machine, où ils ont obtenu des résultats égaux, si ce n'est supérieurs aux réseaux de neurones (Quinlan, 1986, Langley, 1996). En psychologie, ils ont été utilisés pour simuler divers phénomènes, tels que l'apprentissage verbal (Feigenbaum & Simon, 1962, 1984; Hintzman, 1968), l'apprentissage implicite (Ling & Marinov, 1994), l'apprentissage de la langue maternelle (voir plus bas), l'acquisition de l'expertise (voir plus bas), la formation de concepts (Gobet, Richman, Staszewski & Simon, 1997; Fisher, Pazzani & Langley, 1991), et la tâche piagétienne de la balance (Gobet, 1999). Anderson et Bower (1973) ainsi que Barsalou et Bower (1984) ont émis plusieurs critiques à propos de l'emploi des réseaux de discrimination en psychologie; ils notent par exemple que ces modèles accordent trop d'importance à l'ordre avec lequel les attributs sont testés, et qu'ils n'offrent pas de mécanismes permettant de savoir si un stimulus a déjà été perçu auparavant. Ces critiques sont réfutées en détail par Simon (1979, pp. 99-101) ainsi que par Feigenbaum et Simon (1984), articles auxquels nous référons le lecteur pour de plus amples détails.

EPAM

EPAM (Elementary Perceiver And Memorizer) possède une histoire respectable, puisqu'il a été l'un des premiers programmes informatiques à modéliser des processus psychologiques (Feigenbaum & Simon, 1962; 1984). Le but originel d'EPAM était d'offrir un modèle de la manière dont les humains apprennent à discriminer parmi un certain nombre de stimuli et à associer une réponse à chaque stimulus. La théorie inclut un ensemble d'hypothèses en ce qui concerne les structures de mémoire (voir Figure 2), les mécanismes de reconnaissance de l'information, et les règles d'acquisition permettant d'ajouter de nouveaux éléments en mémoire. EPAM est réalisé sous la forme d'un programme d'ordinateur, ce qui permet de dériver des prédictions très précises.

 Insérer Figure 2

La structure de base d'EPAM consiste en un réseau de discrimination, qui permet, nous l'avons vu, d'assigner un stimulus à sa représentation en mémoire. Dans un premier temps, le stimulus est analysé en un ensemble de paires attribut-valeur et/ou de sous-objets.² Le réseau de discrimination consiste en un réseau séquentiel de tests portant sur les caractéristiques des objets. Dans les premières versions d'EPAM, les tests étaient binaires (présence ou absence d'une caractéristique); à partir d'EPAM III, des tests n-aires sont utilisés, de sorte que de multiples branches peuvent provenir d'un même noeud.

Deux types d'apprentissage se produisent avec EPAM. Lorsqu'un objet est reconnu de manière incorrecte, un nouveau test et un nouveau noeud sont ajoutés au réseau; le test porte sur une caractéristique de l'objet permettant de le différencier de l'image avec laquelle il avait été confondu. Il s'agit là du mécanisme de *discrimination*. Si l'objet est reconnu de manière correcte, de nouvelles caractéristiques sont ajoutées à l'image du dernier noeud atteint lors de l'analyse de l'objet. C'est le mécanisme de *familiarisation*. Ces deux types d'apprentissage opèrent de manière discrète. Une des caractéristiques de la théorie est que le processus de localisation d'une trace s'effectue de manière aussi

économique que possible, seules les propriétés nécessaires à l'identification des stimuli étant prises en compte.

EPAM comprend en outre des postulats sur le fonctionnement et les limitations de la MCT et sur les stratégies attentionnelles utilisées par les sujets. Dans les premières versions d'EPAM, la MCT est limitée à un nombre fixe d'éléments. Dans les versions plus récentes (Richman, Staszewski & Simon, 1995), la MCT est divisée en compartiments visuels et auditifs; le compartiment visuel combine un nombre limité de chunks avec une représentation visuo-spatiale et le compartiment auditif combine un nombre limité de chunks avec le concept de la boucle articulatoire (Baddeley & Hitch, 1974).

EPAM a simulé avec succès une grande variété de phénomènes mis en évidence par les recherches sur l'apprentissage verbal, tels que l'oscillation, l'interférence rétroactive, la forme en U de la courbe d'apprentissage, ainsi que les effets de familiarité et de signification des stimuli (Simon & Feigenbaum, 1964; Feigenbaum & Simon, 1984; Gregg & Simon, 1967). Une variante d'EPAM a simulé quelques résultats importants de la psychologie des experts, notamment le type de configurations reconnues dans une position d'échecs (Simon & Gilmarin, 1973). EPAM a également été employé pour simuler l'apprentissage de l'épellation (Simon & Simon, 1973), l'effet de contexte sur la perception de lettres (Richman & Simon, 1989), le comportement d'experts dans la mémorisation de chiffres (Richman, Staszewski & Simon 1995), et les processus de formation de concepts (Richman, 1991; Gobet et al., 1997). On peut donc considérer cette théorie comme un candidat sérieux parmi les nombreux modèles portant sur la mémoire et l'apprentissage.

CHREST

CHREST (Chunk Hierarchy and REtrieval STructures) est une modification d'EPAM qui vise à simuler une grande quantité de données empiriques au sein d'une seule architecture cognitive. En particulier, cette architecture permet d'étudier le rôle du chunking et de la création de schémas dans l'apprentissage de domaines complexes (Gobet et al., 2001). Une attention particulière est donnée au rôle de l'environnement,

dont la structure est partiellement captée par les algorithmes d'apprentissage de l'architecture. Les simulations centrées autour de CHREST concernent avant tout des questions de recherche fondamentale, quoique des applications en éducation, et en particulier dans l'élaboration de didacticiels, soient possibles à long terme (Gobet & Wood, 1999).

Tout comme EPAM, CHREST utilise deux types d'apprentissage (discrimination et familiarisation), et comprend des postulats sur le fonctionnement et les limitations de la MCT ainsi que sur les stratégies attentionnelles utilisées par les sujets. En outre, des méthodes ont été proposées pour augmenter la puissance de calcul du réseau de discrimination de sorte à implémenter un système de production et un réseau sémantique (Gobet, 1996). Ainsi, par rapport à EPAM, la contribution de CHREST est de proposer des mécanismes par lesquels certains chunks (a) sont reliés par des arcs latéraux, et (b) deviennent des structures de données plus complexes, appelées *chablons* ("templates", en anglais, Gobet & Simon, 1996a, 2000). Comme les schémas traditionnels, les chablons comportent une partie fixe et une partie variable. Il est important de noter que malgré la popularité du concept de schéma en psychologie, très peu de mécanismes informatiques ont été proposés dans la littérature pour expliquer leur acquisition (pour une discussion de quelques rares exemples, voir Lane, Gobet & Cheng, 2000).

CHREST a simulé, en général avec succès, divers phénomènes issus de l'apprentissage des catégories syntaxiques (Gobet & Pine, 1997; Jones, Gobet & Pine, 2000b; Croker, Pine & Gobet, 2000), de l'apprentissage du vocabulaire (Jones, Gobet & Pine, 2000a), de l'épreuve Piagétienne de la balance (Gobet, 1999), de la perception et de la mémoire des joueurs d'échecs (De Groot & Gobet, 1996; Gobet, 1993; Gobet & Simon, 1996a, 1996b, 2000; Simon & Gobet, 2000), de l'acquisition de représentations multiples en physique (Lane, Cheng & Gobet, 2000), et de la mémorisation de programmes d'ordinateur (Gobet & Oliver, 2000).

Nous considérons maintenant en détail la manière dont CHREST est employé dans deux domaines distincts: l'acquisition de connaissances par les experts et l'acquisition de la langue maternelle par les enfants. Pour chaque domaine, nous commencerons par une

brève présentation des résultats empiriques principaux et des théories pertinentes, avant de décrire la manière dont CHREST a été utilisé pour rendre compte des données.

Expertise au jeu d'échecs

Le jeu d'échecs a été et est encore l'un des sujets d'étude favoris des psychologues intéressés à l'expertise et à son développement. Les raisons de cet intérêt sont multiples: complexité du domaine, présence d'une échelle numérique classant les joueurs, fertilisation croisée avec l'intelligence artificielle (voir Gobet, 1993, pour une discussion approfondie). Il existe une littérature considérable sur le sujet, qui est discutée en détail par Holding (1985), Saariluoma (1995), et Gobet (1993; 1998b).

Perception et mémoire

Les recherches sur l'expertise aux échecs peuvent être catégorisées en deux lignes principales: travaux sur la mémoire et la perception d'une part, et travaux sur la résolution de problème d'autre part. Plusieurs phénomènes liés à la perception des maîtres d'échecs ont été établis. Par exemple, leurs mouvements oculaires sont plus rapides et plus efficaces que ceux de joueurs faibles, et ils reconnaissent l'identité des pièces et les menaces élémentaires plus rapidement (De Groot & Gobet, 1996; Saariluoma, 1995). Pour ce qui est de la mémoire, il est établi avec suffisamment de certitude que les experts encodent l'information sous forme de chunks, ce qui leur permet d'accéder rapidement de l'information telle que jugements stratégiques sur la position, coups potentiels, menaces possibles, etc. (De Groot, 1978; Chase & Simon, 1973). Les experts contrebalancent la capacité limitée de leur MCT visuo-spatiale, estimée à trois chunks (Gobet & Simon, 2000), par un encodage rapide en MLT (Charness, 1976; Gobet & Simon, 1996a). Des expériences récentes (Saariluoma, 1995; Gobet & Simon, 1996b) indiquent que le nombre de chunks estimé par Chase et Simon (1973) entre 10,000 et 100,000, et considéré comme trop élevé par Holding (1985), est raisonnable, et qu'il semble même sous-estimer le nombre réel. De nombreux résultats (De Groot, 1978; Gruber & Ziegler, 1990; De Groot & Gobet, 1996) suggèrent que les experts aux échecs emploient également des représentations de haut niveau plus complexes que les chunks proposés par

Chase et Simon (1973). Finalement, il a été démontré que, dans les tâches de rappel, les experts maintiennent leur supériorité avec des positions aléatoires, même avec un temps de présentation court (Gobet & Simon, 1996b, 2000).

Outre CHREST, des modèles informatiques symboliques de la perception et de la mémoire aux échecs ont été proposés par Simon et Barenfeld (1969), Simon et Gilmarin (1973), et Saariluoma et Laine (2001). Des modèles non-symboliques ont été développés par Lories (1992), ainsi que par Hyötyniemi et Saariluoma (1998).

Modèle

Le premier domaine d'application de CHREST a été l'expertise au jeu d'échecs. Dans ce domaine, CHREST vise principalement à offrir un modèle informatique expliquant les différences entre novices et experts pour ce qui est de la perception et de la mémoire. Combinant deux modèles antérieurs basés sur les mécanismes de chunking (Simon & Barenfeld, 1969; Simon & Gilmarin, 1973), CHREST incorpore le concept de chablon dans le but de pallier deux faiblesses principales de ces modèles: (a) difficulté d'expliquer les connaissances schématiques de haut niveau; et (b) sur-estimation du temps d'encodage en MLT. Le modèle comprend un certain nombre de paramètres temporels, tels que le temps de créer un nouveau chunk (8 s) ou le temps d'encoder un chunk en MCT (50 ms; cf. Gobet, 1993, et Gobet & Simon, 2000). L'apprentissage—c.-à-d. le développement du réseau de discrimination et la création de liens latéraux—se fait en utilisant un matériel écologique et représentatif de l'environnement: parties de maîtres prises d'une volumineuse banque de données. Finalement, une des contributions de cette recherche est de proposer des mécanismes pour la création de schémas. Ces mécanismes utilisent la différence d'information entre un noeud donné et deux types de noeuds: les descendants de ce noeud, et les noeuds qui lui sont reliés par arcs latéraux. Si le même type d'information (p. ex., dans la Figure 3, Pion blanc et case 'e4') est présent suffisamment de fois (déterminé par un paramètre), une variable est créée pour ce noeud, encodant ce type d'information. Par la suite, CHREST peut employer cette variable pour y encoder de nouvelles valeurs rapidement (250 ms).

Insérer Figure 3

La théorie des chablons, qui est une généralisation de CHREST, rend mieux compte des données empiriques sur l'expertise aux échecs que les théories alternatives (Gobet, 1998b; Simon & Gobet, 2000). Dans cet article, nous nous limiterons aux données qui ont été simulées par CHREST.

Simulations

De Groot et Gobet (1996) décrivent une série de simulations sur les mouvements oculaires d'experts et de débutants lors de la présentation d'une position d'échecs durant cinq secondes. CHREST explique de manière satisfaisante le temps de fixation moyen, la variabilité des temps et le type de cases fixées. Une série de simulations (Gobet, 1993; Gobet & Simon, 1996b; Gobet & Simon, 2000) rend compte du rappel de divers types de positions: positions tirées de parties, positions aléatoires, et positions modifiées par image en miroir. Dans tous ces cas, CHREST reproduit les effets dus au type de position et ceux dus au niveau d'expertise. Gobet et Simon (2000) décrivent en détail l'effet du temps de présentation sur le rappel de positions normales et aléatoires, analysant des variables tels que le pourcentage de rappel, le nombre et le type d'erreurs, ainsi que le nombre et la taille des chunks employés pour reconstituer la position. Gobet (1993) décrit des simulations sur le rappel de positions multiples et l'effet de tâches interférentes sur la mémorisation de positions.

Une caractéristique intéressante de ces simulations est qu'elles ont remis en question certains résultats unanimement acceptés, et que les prédictions du modèle ont été validées. Par exemple, depuis les travaux de Chase et Simon (1973), il avait été généralement accepté que les experts montrent une nette supériorité avec un matériel significatif, mais que leur avantage disparaît totalement avec un matériel aléatoire. Cependant, les simulations menées avec CHREST dans le cas des échecs ont mené à une prédiction différente : comme les experts sont capables de reconnaître, simplement par hasard, quelques chunks dans les positions aléatoires, ils devraient faire preuve d'une supériorité

même avec ce type de matériel, quoique celle-ci devrait être réduite par rapport à un matériel pourvu de sens. Une méta-analyse (Gobet & Simon, 1996c), ainsi que de nouvelles expériences (p. ex., Gobet & Waters, 2000), ont confirmé cette prédiction. La supériorité des experts avec un matériel aléatoire est documentée dans d'autres domaines, tels que la mémoire pour programmes d'ordinateur (Gobet & Oliver, 2000).

Dans la plupart des simulations que nous venons de décrire, l'intérêt s'est centré sur des individus experts, ou du moins au bénéfice d'une expérience importante avec le domaine. Que se passe-t-il avec des joueurs tout au début de leur apprentissage? Gobet (2001) montre comment CHREST explique le comportement de deux sujets possédant des connaissances minimales du jeu d'échecs et entraînés par Saariluoma et Laine (2001) à mémoriser des positions. En particulier, le modèle réplique la courbe typique d'une fonction de puissance ($y = ax^b$) présente dans les données humaines.

Résolution de problème

De Groot (1978) a montré que les experts ne diffèrent que peu des non-experts en ce qui concerne des variables telles que le nombre de coups anticipés ou la structure de la recherche au travers de l'espace de problème. Quel que soit leur niveau, les joueurs font preuve d'une recherche sélective et utilisent un approfondissement progressif, c'est-à-dire qu'ils visitent plusieurs fois la même séquence de coups. En moyenne, les experts ont tendance à anticiper plus de coups en profondeur que les non-experts, quoique la différence soit minime (Holding, 1985; Saariluoma, 1995; Gobet, 1998a). CHREST non compris, des modèles informatiques et psychologiques des processus de recherche aux échecs ont été proposés par Newell et Simon (1972) et par Wagner et Scurrah (1971). En intelligence artificielle, Pitrat (1977) et Wilkins (1980) ont développé des systèmes de production effectuant une recherche sélective (environ 100 positions au maximum), ce qui est psychologiquement plausible.

Modèle et simulations

La nécessité d'un modèle couvrant l'ensemble de l'expertise au jeu d'échecs—combinant perception, mémoire, et résolution de problème—a été mentionnée

à plusieurs reprises dans la littérature, par exemple par Koedinger et Anderson (1990), et l'un des buts de CHREST est de combler cette lacune. Des mécanismes permettant d'employer CHREST pour réaliser un système de production (Gobet & Jansen, 1994; Gobet, 1996; voir Figure 4) ont étendu ce modèle et ont permis quelques explorations dans la résolution de problème aux échecs (Gobet, 1997, 1998a). En particulier, SEARCH (Gobet, 1997), un modèle stochastique basé sur CHREST, incorpore plusieurs idées issues de la recherche sur la résolution de problème aux échecs, et estime, en fonction du nombre de chunks et de chablon, des données comportementales telles que la profondeur ou la vitesse de calcul. Ce modèle abstrait fait une série de prédictions testables sur le développement de débutant à grand maître. Par exemple, selon SEARCH, la profondeur moyenne de calcul suit une loi de puissance ($y = ax^b$), avec le nombre de chunks ou le niveau de jeu comme variables indépendantes, en accord avec les résultats empiriques. Malgré ces progrès, ce qui manque encore est un modèle informatique capable de jouer à un niveau compatible avec son habileté à mémoriser divers types de positions, et capable de simuler les caractéristiques d'anticipation montrées par les humains, y compris le type d'erreurs qu'ils commettent.

 Insérer Figure 4

Finalement, CHREST explique plusieurs résultats empiriques dans d'autres domaines d'expertise (Richman, Gobet, Staszewski & Simon, 1996; Simon & Gobet, 2000). En particulier, des simulations ont été conduites en ce qui concerne la mémorisation de programmes d'ordinateur (Gobet & Oliver, 2000) et la manière dont plusieurs représentations peuvent être combinées en physique (Lane, Cheng & Gobet, 2000).

Apprentissage de la langue maternelle

Nous tournons maintenant notre attention vers le second domaine sous investigation, l'apprentissage de la langue maternelle. L'intérêt se centrera sur

l'acquisition des catégories syntaxiques, avec une excursion dans le domaine de l'acquisition du vocabulaire.

La manière dont les enfants apprennent à reconnaître et à employer les catégories syntaxiques telles que nom et verbe est une question essentielle de l'étude scientifique de la langue. Cependant, bien que des progrès non-négligeables aient été effectués par les théories d'obédience innéiste (Pinker, 1984; Wexler, 1994) et les approches basées sur les réseaux de neurones (Elman, 1993; Rohde & Plaut, 1999), il n'existe actuellement aucun modèle suffisamment détaillé pour permettre des prédictions empiriques précises sur le développement de la syntaxe.

Les modèles de la famille EPAM/CHREST—tout comme les réseaux de neurones, d'ailleurs—font partie d'un ensemble d'approches informatiques basées sur l'analyse distributionnelle de l'environnement (dans ce cas, le langage), approches dont on a pu récemment observer une résurgence. Ces approches ont une longue histoire en psychologie, et ont montré récemment l'énorme quantité d'information linguistique qui est présente dans de grands corpus de textes écrits et oraux (Finch & Chater, 1992; Cartwright & Brent, 1997). Cependant, ces théories ont également été critiquées pour plusieurs raisons, en particulier du fait que certaines propriétés du langage, telles que la dépendance éloignée, ne peuvent pas être acquises par une analyse distributionnelle, ou que les mécanismes proposés ne sont pas plausibles psychologiquement (Pinker, 1984). Il faut toutefois noter que les travaux d'Elman (1993), utilisant des langages artificiels, ont établi que de telles réfutations *a priori* ne sont pas tenables.

Ilots verbaux

Parmi les approches non-informatiques, une des explications les plus influentes du développement grammatical est l'hypothèse constructiviste récemment proposée par Tomasello (1992) d'un *îlot verbal*. Selon cette hypothèse, les enfants commencent la phase de production multi-mots sans la connaissance de catégories syntaxiques telles que "Sujet" ou "Objet direct" ou de catégories sémantiques telles que "Agent" et "Patient", et construisent progressivement des catégories spécifiques à un verbe donné (telles que, en anglais, "Hitter" et "Hittee", pour le verbe "Hit") sur la base de leur expérience avec des

prédicats particuliers. Les premières grammaires des enfants peuvent alors être vues comme des inventaires de structures prédicatives lexicalement spécifiques (ou “îlots verbaux”) acquises au travers d’un processus d’analyse distributionnelle, et l’acquisition par les enfants de marquage syntaxique et morphologique est supposée progresser de verbe en verbe. Tout se passe comme si la connaissance des catégories syntaxiques, d’abord limitée à des îlots déconnectés, se construit progressivement en reliant ces îlots pour former une structure intégrée.

Tomasello propose que les enfants débudent la construction de catégories paradigmatisées telles que nom et verbe seulement lorsqu’ils commencent à opérer sur des exemples de ces catégories en tant qu’arguments de prédicats. L’implication est que, durant l’apprentissage d’une langue comme l’anglais, les enfants vont former la catégorie de nom relativement rapidement, car les noms sont employés comme arguments de prédicats tôt dans le développement. Par contre, la formation de la catégorie paradigmatisée de verbe ne se produira que plus tard lorsque les enfants commencent à employer les verbes comme arguments d’autres prédicats, par exemple dans des constructions comprenant deux verbes, telles que “Want to + *Verbe*” or “Can’t + *Verbe*”.

L’hypothèse de l’îlot verbal explique un certain nombre de phénomènes propres au début du langage multi-mots. Elle peut expliquer la manière dont, dans la petite enfance, les verbes sont employés de manière lexicalement spécifique. Cette hypothèse peut également expliquer la nature limitée qui caractérise les règles d’emploi de l’ordre des mots dans la petite enfance. Par exemple, Akhtar et Tomasello (1997) ont montré que les jeunes enfants non seulement ne généralisent pas leur connaissance de l’ordre des mots *Sujet-Verbe-Objet* d’un verbe à un autre, mais aussi qu’ils sont incapables d’employer cet ordre comme un indice facilitant la compréhension de phrases comprenant de nouveaux verbes. Finalement, cette hypothèse explique les différences dans la flexibilité avec laquelle les enfants emploient les noms et les verbes au début de leur production multi-mots, les jeunes enfants incorporant facilement de nouveaux *noms* dans des structures de verbes familières, mais non l’inverse (Akhtar & Tomasello, 1997).

Bien que l'hypothèse d'îlot verbal rende compte des données développementales de manière raisonnable, elle a deux faiblesses importantes. D'une part, les verbes sont définis au sein de la théorie comme "anything in the child's language that functions as a predicate" (Tomasello, 1992, p. 11). Cela suggère que les phénomènes liés à l'îlot verbal reflètent une relation plus ou moins directe entre la distinction nom-verbe au niveau syntaxique et la distinction argument-prédicat au niveau sémantico-pragmatique. Toutefois, comme Maratsos (1990) l'a montré, la corrélation entre ces distinctions est loin d'être parfaite, même en anglais, ce qui suggère qu'il n'est pas satisfaisant de se référer à une telle corrélation pour expliquer les phénomènes d'îlot verbal chez les enfants.

La seconde faiblesse est que certains aspects du langage enfantin multi-mots ne sont pas compatibles avec une version stricte de cette hypothèse. Par exemple, Pine, Lieven et Rowland (1998) ont montré que de nombreux enfants acquièrent des structures lexicalement spécifiques basées autour de mots certes très fréquents, mais qui ne sont pas normalement définis en tant que prédicats (p. ex., pronoms marqués au nominatif tels que "I" et "He" et noms propres tels que "Mummy" et "Anne"). De plus, ces "îlots" de pronoms et de noms propres semblent fonctionner de manière à structurer des éléments du langage enfantin pour lesquels des verbes servent comme arguments. Ces données suggèrent que la spécificité lexicale du langage multi-mots de la petite enfance n'est pas toujours une spécificité lexicale propre aux verbes, ou même propre aux prédicats en tant que tels. En conséquence, ces effets d'îlots verbaux semblent être simplement un cas spécial d'effets généraux liés à la fréquence sur l'acquisition par les enfants de structures lexicalement spécifiques.

Modèle

Dans le cadre de l'acquisition de la syntaxe, Croker et al. (2000), Gobet et Pine (1997), ainsi que Jones et al. (2000b) discutent en détail une variante de CHREST nommée MOSAIC (Model Of Syntax Acquisition In Children). Les postulats principaux du modèle sont que: (a) les catégories syntaxiques sont activement construites par l'enfant qui utilise un apprentissage portant sur les propriétés statistiques de l'environnement; et

(b) cet apprentissage est limité par des contraintes cognitives telles que vitesse d'apprentissage et capacité de la MCT.

Les mécanismes essentiels du modèle peuvent être décrits de la manière suivante: MOSAIC utilise les énoncés de l'input, codés orthographiquement, pour créer un réseau de discrimination. Dans la version actuelle, l'input consiste en un corpus d'énoncés naturels produits par une mère interagissant avec son enfant. Les noeuds qui sont suffisamment similaires en ce qui concerne les tests qu'ils possèdent sont reliés par des arcs latéraux, créant ainsi un réseau sémantique. Durant la production d'énoncés, le programme peut ou bien utiliser les tests en dessous d'un noeud (output par reproduction) ou utiliser les arcs latéraux (output par génération).

Création d'arcs latéraux

MOSAIC crée des arcs latéraux qui peuvent relier deux noeuds (ces noeuds peuvent se trouver à n'importe quel niveau du réseau). Les arcs latéraux combinent des noeuds qui sont employés dans le même contexte, où *contexte* est défini comme l'ensemble des tests qui suivent immédiatement un noeud donné. La manière dont MOSAIC crée les noeuds et les tests qui leur sont associés implique que les tests qui se trouvent en dessous d'un noeud donné comportent des mots qui, dans l'input, ont suivi les mots contenus par le noeud en question. Par exemple, dans la Figure 5, les mots "moves", "sits", "walks", et "chases" doivent avoir suivi "cat"; en d'autres termes, la séquence "cat sits" doit avoir été présente.

 Insérer Figure 5

Lorsque deux noeuds partagent suffisamment de tests, c.-à-d. que le contexte dans lequel ces deux noeuds sont employés est similaire, ils sont reliés par un arc latéral. Le nombre minimal de tests qui doivent être communs est déterminé par un paramètre, dont la valeur est fixée arbitrairement dans les simulations (en général, 20). Par exemple, dans la Figure 5, "cat" et "dog" seraient reliés par un arc latéral si ce paramètre avait la valeur

3—trois des noeuds en dessous du noeud “cat” coïncident avec trois des noeuds en dessous du noeud “dog”.

Production d'énoncés à partir du réseau

Le modèle produit des énoncés en commençant par le noeud-racine et en traversant le réseau jusqu'à ce qu'un noeud terminal soit rencontré. En traversant le réseau, les arcs suivant un test ou les arcs latéraux peuvent être utilisés. Si la production de l'énoncé s'effectue verticalement en n'employant que des arcs suivant un test (output par reproduction), cela signifie que l'énoncé était présent dans l'input utilisé durant l'apprentissage. Par contre, si un arc latéral est employé (output par génération), il est possible que l'énoncé soit nouveau, c'est-à-dire qu'il n'est pas présent dans le corpus. Un exemple clarifiera le mécanisme de génération. Dans le réseau de la Figure 5, l'énoncé pourrait commencer avec “cat”, suivre l'arc latéral menant à “dog”, et ensuite continuer l'énoncé avec n'importe quelle séquence de noeud en dessous de “dog”. Ce mécanisme produit des énoncés qui ne sont pas présents dans l'input, tels que “cat runs” et “dog moves”.

Simulations

En général, la méthode de simulation consiste (a) à développer un réseau de discrimination (incluant des arcs latéraux), en utilisant comme input un corpus d'énoncés naturels produits par des parents; (b) à produire un grand nombre d'énoncés à partir du modèle (plusieurs centaines de milliers); (c) à échantillonner aléatoirement un nombre d'énoncés semblable à celui obtenu dans les échantillons des sujets enfants; (d) à analyser les énoncés produits par MOSAIC de la même manière que les énoncés humains; et (e) à comparer quantitativement les énoncés humains avec ceux produits par MOSAIC. L'échantillonnage lors de la phase (c) se fait ou bien en utilisant le nombre absolu d'énoncés dans l'échantillon à simuler, ou bien en utilisant le nombre d'énoncés qui possèdent une propriété donnée (p. ex., énoncés avec une erreur dans l'emploi de pronoms).

Le programme explique les phénomènes empiriques liés à la présence d'îlots verbaux dans le début du langage multi-mots enfantin (Jones et al., 2000b). Comme proposé par Tomasello, (1992), de nombreux îlots sont présents avec les *verbes*, mais peu sont présents avec des *noms communs*. De plus, MOSAIC simule la présence d'*îlots non-verbaux*, tels qu'*îlots de pronoms* (p. ex., "I" ou "He") et *îlots de noms propres* (p. ex., "Mummy", "Daddy", ou le prénom de l'enfant). En particulier, les îlots de pronoms sont fréquents chez les enfants, et, du fait que les pronoms prennent des verbes comme arguments, ces îlots posent problème pour la version stricte de la théorie de Tomasello.

Finalement, MOSAIC explique de manière mécanique et économique pourquoi certains items lexicaux fonctionnent comme îlots dans le langage des enfants, et pourquoi d'autres ne le font pas: la création des îlots reflète la structure statistique de l'environnement. Cependant, il est important de noter que MOSAIC fait plus que simplement enregistrer les propriétés distributionnelles de l'input, car sa production est plus proche de celle des enfants que de celle de leur mère. De fait, c'est bien la combinaison des contraintes imposées à MOSAIC (p. ex., apprentissage d'un seul mot à la fois) et de la fréquence d'apparition des énoncés dans l'input qui permet à MOSAIC de répliquer les données empiriques provenant des enfants.

Erreurs de type "infinitif optionnel" et de marquage de cas

Il est typique que les enfants anglophones entre deux et trois ans commettent des erreurs tels que "He do it" et "She doing it" (au lieu de "He does it" et "She does it"). Plusieurs théories ont été proposées pour expliquer ces erreurs, parmi lesquelles l'une des plus influentes est celle de l' "infinitif optionnel" (Wexler, 1994). Cette théorie innéiste propose qu'il existe une étape dans le développement du langage durant laquelle les enfants ont fixé correctement tous les paramètres de base liés à l'inflexion et à la structure de phrase, mais qu'ils ne possèdent pas encore la connaissance que le temps est obligatoire avec les phrases contenant un élément verbal fini. Cette théorie fait donc l'hypothèse que les enfants possèdent très tôt une grande quantité de connaissances abstraites.

Simulations

Une explication alternative, défendue par Croker et al. (2000), est que la formation de relations syntaxiques dans le langage enfantin est largement déterminée par l'input verbal reçu par l'enfant. Dès lors, la distribution statistique de l'environnement linguistique, en grande partie influencée par les productions verbales des parents, va avoir un impact direct sur les formes lexicales que l'enfant va employer, avec la conséquence que la plupart des patterns produits par l'enfant vont avoir des modèles dans le langage adulte. Par exemple, l'énoncé "she going", qui est considéré par Wexler (1994) comme étant dû à une omission du marquage de temps, proviendrait d'énoncés présents dans l'input tels que "where is she going?" ou "is she going to the shops?" De même, des énoncés tels que "look at him go" constituent un modèle pour des erreurs de marquage de cas tels que "him go". Par contre, les types d'énoncés interdits par l'hypothèse de l'infinitif optionnel, tels que "it are", ne possèdent pas de modèles dans le langage parental.

Bien que MOSAIC emploie des mécanismes d'apprentissage très simples, ce modèle est capable de reproduire toutes les erreurs décrites dans la littérature pour ce qui est de l'anglais (Harris & Wexler, 1996; Schutze & Wexler, 1996), y compris celles qui ne sont pas expliquées par l'hypothèse de l'infinitif optionnel. De plus, le modèle rend compte de phénomènes similaires en hollandais (Freudenthal et al., sous presse). MOSAIC produit certaines de ces erreurs par simple reproduction de fragments présents dans l'input (voir exemples ci-dessus). D'autres erreurs sont produites par le mécanisme de génération. Par exemple, l'erreur "her does" est produite du fait que MOSAIC a créé un arc latéral entre les noeuds "her" et "that". Au total, ce modèle, dans lequel des mécanismes limités d'analyse distributionnelle interagissent avec les propriétés statistiques de l'input, offre une bonne approximation des données empiriques.

Acquisition du vocabulaire

Une simplification de MOSAIC est que les énoncés de l'input sont encodés orthographiquement. Jones et al. (2000a) montrent que les mécanismes de base du modèle permettent également de traiter des énoncés encodés de manière phonétique.

Employant les mêmes corpus que pour la simulation de l'acquisition de catégories syntaxiques, ces auteurs décrivent des simulations portant sur l'acquisition du vocabulaire.

Le modèle combine MOSAIC avec l'idée de boucle phonologique (Baddeley & Hitch, 1974) et avec certains paramètres temporels liés au traitement de chunks en MCT (Zhang & Simon, 1985). Comme avec l'acquisition de catégories syntaxiques, l'apprentissage s'opère par la création d'un réseau de discrimination, et l'interaction entre la MLT et la boucle phonologique se fait au travers du concept de chunk.

Jones et al. (2000a) montrent que le modèle explique de manière raisonnable la performance d'enfants dans la tâche de *répétition de non-mots*, une tâche qui est souvent présentée comme un diagnostic puissant de l'apprentissage du vocabulaire. Leurs simulations portent sur les expériences princeps de la répétition de non-mots: l'étude de Gathercole et Baddeley (1989) avec des enfants de 4 et 5 ans, et l'étude de Gathercole et Adams (1993) avec des enfants de 2 et 3 ans.

Discussion

Dans cet article, nous avons montré comment des principes similaires expliquent de nombreuses données dans les domaines de l'acquisition de l'expertise et de l'apprentissage de la langue maternelle. Dans tous ces cas, des modèles informatiques reproduisant les phénomènes clés ont été présentés.

Evaluation des simulations

La perception et la mémoire aux échecs, qui ont été les premiers domaines d'application de CHREST, sont également les domaines les mieux développés. Un nombre élevé de données empiriques sont simulées en détail, et, dans les cas où aucune simulation n'a été effectuée, une application informelle du modèle explique la plupart des données (Gobet, 1998b)—évidemment avec moins de précision. Des progrès ont également été enregistrés avec la résolution de problème et avec la généralisation à d'autres domaines d'expertise, tels que la physique et la programmation.

MOSAIC est capable de simuler un certain nombre de phénomènes liés à l'acquisition de la syntaxe, tels que phénomènes d'îlot verbal et d'infinitif optionnel. A notre connaissance, aucun autre modèle informatique n'a été appliqué à l'ensemble de ces données. Un avantage offert par MOSAIC est que cette approche permet de développer des simulations basées sur de grands corpus de données naturelles, et non seulement sur des descriptions abstraites de ces données. De manière intéressante, le modèle ne rend pas compte des données quantitatives aussi bien que dans le cas de l'expertise aux échecs, probablement du fait que l'input employé pour la création des réseaux, bien que provenant de l'un des corpus les plus grands actuellement, ne représente qu'une fraction de l'input reçu par un enfant. Un corpus plus dense semble nécessaire pour expliquer les données empiriques, en particulier si les simulations visent à répliquer les différences individuelles.

Pour ce qui est de l'acquisition du vocabulaire, et bien que les simulations soient plus limitées, on peut mentionner une contribution importante. Cette recherche, qui a combiné CHREST avec l'approche de la mémoire de travail, a surmonté une sérieuse limite de la théorie proposée par Baddeley, Gathercole et Papagno (1998): le modèle propose des mécanismes par lesquels de nouveaux mots peuvent être appris en combinant des séquences de sons et par lesquels le vocabulaire existant influence l'acquisition de nouveaux mots, ce qui manque totalement dans la théorie de Baddeley et al.

Il est important de noter que les simulations de chaque domaine n'ont pas été effectuées en isolation, mais qu'elles se sont répercutées dans les autres domaines, au travers de changements théoriques aussi bien que par la création de bibliothèques de programmes réutilisables dans d'autres projets. Ainsi, plusieurs programmes implémentant CHREST dans le cadre de l'expertise aux échecs ont été utilisés avec des changements minimes pour simuler des données sur le rôle de représentations multiples en physique (Lane, Cheng & Gobet, 2000) et sur la mémoire de programmes d'ordinateurs (Gobet & Oliver, 2000). Pareillement, les programmes implémentant MOSAIC ont été utilisés pour la modélisation de l'acquisition du vocabulaire.

Malgré ces résultats encourageants, il reste évidemment beaucoup à faire. Quelques exemples suffiront. Avec CHREST, il serait souhaitable d'une part de développer un modèle unifié dans un domaine particulier d'expertise, et, d'autre part, de conduire des simulations dans d'autres domaines riches en données empiriques, tels que la médecine ou la musique. Avec MOSAIC, il serait intéressant de simuler des données portant sur l'apprentissage de la syntaxe avec des langues germaniques ou romanes, et de modéliser les différences individuelles. Finalement, on pourrait appliquer le modèle de la répétition de mots à la simulation de l'accélération brusque de l'acquisition du vocabulaire observée autour de 18 mois et aux différentes dynamiques montrées par l'acquisition des noms et des verbes. Il serait également souhaitable de combiner ce modèle avec celui de l'acquisition de la syntaxe; la conséquence serait qu'un seul modèle, apprenant à partir d'énoncés codés sous forme de phonèmes, simulerait des données provenant de l'acquisition du vocabulaire et de l'acquisition de catégories syntaxiques.

Avantages et désavantages offerts par les réseaux de discrimination

Les résultats présentés dans cet article sont la résultante de plusieurs caractéristiques propres aux réseaux de discrimination et de certains postulats additionnels. Comme nous l'avons noté auparavant, les réseaux de discrimination constituent des systèmes auto-organiseurs très sensibles aux propriétés de l'environnement et capables d'apprendre sans supervision. Ces propriétés ont souvent été proposées comme essentielles pour l'étude de la cognition (p. ex., Pfeifer & Scheier, 1999). Nous avons également montré comment les réseaux de discrimination pouvaient constituer les fondations sur lesquelles des systèmes de production et des réseaux sémantiques pouvaient être bâtis. Finalement, ce type de formalisme offre quelques avantages pratiques non-négligeables: transparence, généralisabilité, rapidité d'apprentissage, et réutilisabilité.

De manière importante, certains postulats indépendants du formalisme utilisé jouent également un rôle important, en particulier pour l'acquisition de l'expertise et du vocabulaire; parmi eux, on peut mentionner l'hypothèse que la MCT est très limitée, que les processus cognitifs exigent un temps fixe, et que l'accès en MLT est relativement lent.

Plusieurs désavantages de ce formalisme peuvent être mentionnés. Le premier est que, du fait de l'importance accordée à l'apprentissage, il est nécessaire de posséder un input suffisamment représentatif du domaine à apprendre, aussi bien qualitativement que quantitativement. Cela peut être difficile pour certains comportements ou exiger un input plus riche que ce qui est actuellement à disposition; par exemple, comme nous l'avons mentionné avec MOSAIC, une limitation des simulations sur l'apprentissage de la langue est que l'input à disposition est de taille insuffisante. Un deuxième inconvénient est que l'apprentissage peut produire des réseaux de taille importante (jusqu'à de 300'000 noeuds, dans le cas des échecs), avec de nombreux noeuds reliés par des liens latéraux. Bien qu'une telle organisation des connaissances soit plus plausible, par exemple, qu'une liste bien ordonnée de règles de production, elle soulève des difficultés pour l'analyse de ce que le système a appris. Troisièmement, et c'est une conséquence de cette organisation des connaissances, les chunks appris par CHREST possèdent un haut niveau de redondance (la même information est encodée avec des variations mineures par divers chunks), ce qui, bien que psychologiquement intéressant, peut rendre assez complexe l'analyse du système.

Le quatrième point touche à la flexibilité des modèles. Avec la plupart des systèmes de production "classiques", tels que Soar (Newell, 1990) ou ACT-R (Anderson, 1993), qui sont modulaires, il est possible d'ajouter ou de soustraire des productions sans affecter la structure générale du système. Cela n'est pas le cas des réseaux de discrimination présentés dans cet article. Par exemple, l'addition d'un noeud peut exiger l'addition d'autres noeuds pour l'indexer. De plus, un système comme ACT-R permet un haut niveau de généralisation au sein de ses productions, par le truchement de variables, alors que dans un système comme CHREST, la présence de variables est le produit d'un long apprentissage. A noter cependant que les systèmes de production classiques paient un prix théorique pour cette flexibilité: ils contiennent davantage de paramètres libres.

Enfin, et c'est un problème partagé par tous les systèmes de modélisation, il existe une liberté importante dans le choix des primitives. Dans les simulations présentées dans cet article, ces primitives étaient relativement naturelles (p. ex., pièces et cases dans le cas

du jeu d'échecs, et mots dans le cas de l'apprentissage de la syntaxe), mais il est clair que d'autres représentations étaient possibles.

Une tâche importante pour le futur sera de comparer en détail les réseaux de discrimination avec d'autres formalismes courants en psychologie, tels que réseaux de neurones et systèmes de production. Nous nous sommes limités ici à de brèves remarques sur le rapport entre réseaux de discrimination et systèmes de production, car, comme nous l'avons vu plus haut, les premiers peuvent être utilisés pour construire des règles de production.³ Nous ajouterons deux points. Avec la plupart des systèmes de production adaptatifs, tels que Soar ou ACT-R, l'apprentissage est déterminé par la présence d'impasses dans le processus de résolution de problème; avec les réseaux de discrimination, l'apprentissage est continu. Deuxièmement, les systèmes de production sont orientés par un but (top-down), alors que les réseaux de discrimination sont orientés par les inputs perceptifs (bottom-up).

Conclusion

A plusieurs reprises, Newell a critiqué la tendance des psychologues à développer des micro-modèles, et a souligné la nécessité d'avoir recours à des architectures cognitives (Newell, 1990). Ces dernières ont d'après lui deux caractéristiques essentielles. D'une part, elles doivent être unifiées, c'est-à-dire que leurs principes et mécanismes doivent s'appliquer à l'ensemble de la cognition. D'autre part, elles doivent être informatiques, car, étant donnée la complexité de la cognition humaine, une réalisation sous forme de programme d'ordinateur est nécessaire pour permettre des prédictions précises. L'idée d'architecture cognitive est actuellement réalisée dans des systèmes tels que Soar (Newell, 1990) et ACT-R (Anderson, 1993).

Dans cet article, nous avons illustré comment les réseaux de discrimination ont été employés avec succès dans toute une série de domaines. Nous avons en particulier décrit les recherches centrées autour d'EPAM/CHREST, et avons montré comment elles ont conduit à l'élaboration d'une architecture cognitive basée sur la construction d'arbres de discrimination ainsi que sur la construction de réseaux sémantiques et de systèmes de production à partir de ces arbres. Suivant les exemples offerts par Soar et ACT-R,

EPAM/CHREST explique, au sein d'une même architecture, des phénomènes provenant de l'apprentissage perceptif, de la mémoire, de la formation de concepts, de la résolution de problème, et du langage. Bien que cette architecture ne soit pas sans désavantages, elle constitue, en compagnie d'autres approches basées sur les arbres de discrimination, un outil théorique intéressant pour les psychologues.

Références

- Akhtar, N. & Tomasello, M. (1997). Young children's productivity with word order and verb morphology. *Developmental Psychology*, 33, 952-965.
- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, DC: Winston.
- Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105, 158-173.
- Baddeley, A. D. & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation*, (Vol. 8) (pp. 47-90). New York: Academic Press.
- Barsalou, L. W. & Bower, G. H. (1984). Discriminations nets as psychological models. *Cognitive Science*, 8, 1-26.
- Cartwright, T. A. & Brent, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition*, 63, 121-170.
- Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 641-653.
- Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.) *Visual information processing* (pp. 215-281). New York: Academic Press.
- Crocker, S., Pine, J. M., & Gobet, F. (2000). Modelling the optional infinitive phenomena: A computational account of tense optionality in children's speech. In N. Taatgen & J. Aasman (Eds.), *Proceedings of the Third International Conference on Cognitive Modelling* (pp. 142-149). Veenendaal, The Netherlands: Universal Press.
- De Groot, A. D. (1978). *Thought and choice in chess*. The Hague, Mouton Publishers.
- De Groot, A. D. & Gobet, F. (1996). *Perception and memory in chess: Studies in the heuristics of the professional eye*. Assen: Van Gorcum.
- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48, 71-99.

- Feigenbaum, E. A., & Simon, H. A. (1962). A theory of the serial position effect. *British Journal of Psychology*, 53, 307-320.
- Feigenbaum, E. A., & Simon, H. A. (1984). EPAM-like models of recognition and learning. *Cognitive Science*, 8, 305-336.
- Finch, S. & Chater, N. (1992). Bootstrapping syntactic categories. *Proceedings of the 14th Annual Conference of the Cognitive Science Society* (pp. 820-825). Hillsdale, NJ: Erlbaum.
- Fisher, D. H., Pazzani, M. J., & Langley, P. (Eds.) (1991). *Concept formation: Knowledge and experience in unsupervised learning*. San Mateo: Kaufmann.
- Freudenthal, D., Pine, J. M., & Gobet, F. (2001). Modelling the optional infinitive stage in MOSAIC: A generalisation to Dutch. *Proceedings of the 4th International Conference on Cognitive Modelling*.
- Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, 28, 200-213.
- Gathercole, S. E., & Adams, A.-M. (1993). Phonological working memory in very young children. *Developmental Psychology*, 29, 770-778.
- Gobet, F. (1993). *Les mémoires d'un joueur d'échecs*. Fribourg: Editions Universitaires.
- Gobet, F. (1996). Discrimination nets, production systems and semantic networks: Elements of a unified framework. *Proceedings of the 2nd International Conference on the Learning Sciences*. (pp. 398-403). Evanston IL: Northwestern University.
- Gobet, F. (1997). Roles of pattern recognition and search in expert problem solving. *Thinking and Reasoning*, 3, 291-313.
- Gobet, F. (1998a). Chess players' thinking revisited. *Swiss Journal of Psychology*, 57, 18-32.
- Gobet, F. (1998b). Expert memory: A comparison of four theories. *Cognition*, 66, 115-152.

- Gobet, F. (1999). Simulations of stagewise development with a symbolic architecture. In J. P. Dauwalder & W. Tschacher (Eds.), *Dynamics, synergetics, and autonomous agents* (pp. 143-156). Singapore: World Scientific.
- Gobet, F. (2001). Chunk hierarchies and retrieval structures. Comments on Saariluoma and Laine. *Scandinavian Journal of Psychology*, *42*, 149-155.
- Gobet, F., & Jansen, P. (1994). Towards a chess program based on a model of human memory. In H. J. van den Herik, I. S. Herschberg, & J. E. Uiterwijk (Eds.), *Advances in Computer Chess 7* (pp. 35-60). Maastricht: University of Limburg Press.
- Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C.-H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, *5*, 236-243.
- Gobet, F. & Oliver, I. (2000). Implicit memory for computer programs. (Soumis pour publication).
- Gobet, F., & Pine, J. (1997). Modelling the acquisition of syntactic categories. *Proceedings of the 19th Annual Meeting of the Cognitive Science Society*, p. 265-270. Hillsdale, NJ: Erlbaum.
- Gobet, F., Richman, H. B., Staszewski, J., & Simon, H. A. (1997). Goals, representations, and strategies in a concept attainment task: The EPAM model. *The Psychology of Learning and Motivation*, *37*, 265-290.
- Gobet, F. & Simon, H. A. (1996a). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, *31*, 1-40.
- Gobet, F. & Simon, H. A. (1996b). Recall of random and distorted positions: Implications for the theory of expertise. *Memory & Cognition*, *24*, 493-503.
- Gobet, F., & Simon, H. A. (1996c). Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bulletin & Review*, *3*, 159-163.
- Gobet, F. & Simon, H. A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science*, *24*, 651-682.
- Gobet, F., & Waters, A. (2000). Role of constraints in expert memory. (Soumis pour publication).

- Gobet, F., & Wood, D. J. (1999). Expertise models of learning and computer-based tutoring. *Computers and Education*, *33*, 189-207.
- Gregg, L. W., & Simon, H. A. (1967). An information-processing explanation of one-trial and incremental learning. *Journal of Verbal Learning and Verbal Behavior*, *1967*, 780-787.
- Gruber, H., & Ziegler, A. (1990). Expertisegrad und Wissensbasis. Eine Untersuchung bei Schachspielern. *Psychologische Beiträge*, *32*, 163-185.
- Harris, T. & Wexler, K. (1996) The optional-infinitive stage in child English. In H. Clahsen (Ed.) *Generative perspectives on language acquisition* (pp. 1-42). Philadelphia: John Benjamins.
- Hintzman, D. L. (1968). Explorations with a discrimination net model for paired-associate learning. *Journal of Mathematical Psychology*, *5*, 123-162.
- Holding, D. H. (1985). *The psychology of chess skill*. Hillsdale, NJ: Erlbaum.
- Hyötyniemi, H., & Saariluoma, P. (1998). Simulating chess players' recall: How many chunks and what kind can they be? *Proceedings of the Second European Conference on Cognitive Modelling* (p. 195-196). Nottingham: University Press.
- Jones, G., Gobet, F., & Pine, J. M. (2000a). Learning novel sound patterns. In N. Taatgen & J. Aasman (Eds.), *Proceedings of the Third International Conference on Cognitive Modelling* (pp. 169-176). Veenendaal, The Netherlands: Universal Press.
- Jones, G., Gobet, F., & Pine, J. M. (2000b). A process model of children's early verb use, *Proceedings of the Twenty Second Annual Meeting of the Cognitive Science Society* (pp. 723-728). Philadelphia, USA.
- Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science*, *14*, 511-550.
- Lane, P. C. R., Cheng, P. C.-H., & Gobet, F. (2000). CHREST+: A simulation of how humans learn to solve problems using diagrams. *AISB Quarterly*, *103*, 24-30.
- Lane, P. C. R., Gobet, F., & Cheng, P. C.-H. (2000). Learning-based constraints on schemata, *Proceedings of the Twenty Second Annual Meeting of the Cognitive Science Society* (pp. 776-781). Philadelphia, USA.

- Langley, P. (1996). *Elements of machine learning*. San Francisco, CA: Morgan.
- Levy, S. (1992). *Artificial life: A report from the frontier where computers meet biology*. New York: Vintage Books.
- Ling, C. X., & Marinov, M. (1994). A symbolic model of the nonconscious acquisition of information. *Cognitive Science*, *18*, 595-621.
- Lories, G. (1992, July). *Using a neural network to pre-process chess positions*. Paper presented at the XXVth International Congress of Psychology, Brussels.
- Maratsos, M. P. (1990). Are actions to verbs as objects are to nouns? On the different semantic bases of form, class, and category. *Linguistics*, *28*, 1351-1379.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Pfeifer, R., & Scheier, C. (1999). *Understanding intelligence*. Cambridge: MIT Press.
- Pine, J. M., Lieven, E. V. M., & Rowland, C. F. (1998). Comparing different models of the development of the English verb category. *Linguistics*, *36*, 807-830.
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Pitrat, J. (1977). A chess combinations program which uses plans. *Artificial Intelligence*, *8*, 1977.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*, 81-106.
- Richman, H. B. (1991). Discrimination net models of concept formation. In D. H. Fisher, M. J. Pazzani, & P. Langley (Eds.), *Concept formation: Knowledge and experience in unsupervised learning*, 103-126. San Mateo: Kaufmann.
- Richman, H. B., Gobet, F., Staszewski, J. J., & Simon, H. A. (1996). Perceptual and memory processes in the acquisition of expert performance: The EPAM model. In K. A. Ericsson (Ed.), *The road to excellence* (pp. 167-187). Mahwah: Erlbaum.
- Richman, H. B., & Simon, H. A. (1989). Context effects in letter perception: Comparison of two theories. *Psychological Review*, *96*, 417-432.

- Richman, H. B., Staszewski, J., & Simon, H. A. (1995). Simulation of expert memory with EPAM IV. *Psychological Review*, 102, 305-330.
- Rohde, D. L. T., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, 72, 67-109.
- Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the micro-structure of cognition. Volume 2* (pp. 216-271). Cambridge, MA: MIT Press.
- Saariluoma, P. (1995). *Chess players' thinking*. London: Routledge.
- Saariluoma, P., & Laine, T. (2001). Novice construction of chess memory. *Scandinavian Journal of Psychology*, 42, 137-146.
- Schutze, C. & Wexler, K. (1996). Subject case licensing and English root infinitives. In A. Stringfellow, D. Cahma-Amitay, E. Hughes & A. Zukowski (Eds.), *Proceedings of the 20th Annual Boston University Conference on Language Development*. (pp. 670-681). Somerville, MA: Cascadilla Press.
- Simon, D. P. & Simon, H. A. (1973). Alternative uses of phonemic information in spelling. *Review of Educational Research*, 43, 115-37.
- Simon, H. A. (1979). *Models of Thought*. New Haven: Yale University Press.
- Simon, H. A., & Barenfeld, M. (1969). Information processing analysis of perceptual processes in problem solving. *Psychological Review*, 76, 473-483.
- Simon, H. A., & Feigenbaum, E. A. (1964). An information processing theory of some effects of similarity, familiarity, and meaningfulness in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, 3, 385-396.
- Simon, H. A., & Gilmarin, K. J. (1973). A simulation of memory for chess positions. *Cognitive Psychology*, 5, 29-46.
- Simon, H. A., & Gobet, F. (2000). Expertise effects in memory recall: Comments on Vicente an Wang (1998). *Psychological Review*, 107, 593-600.
- Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge: Cambridge University Press.

- Wagner, D. A. & Scurrah, M. J. (1971). Some characteristics of human problem-solving in chess, *Cognitive Psychology*, 2, 454-478.
- Wexler, K. (1994) Optional infinitives, head movement and the economy of derivations in child grammar. In D. Lightfoot & N. Hornstein (Eds.) *Verb movement* (pp. 305-350). Cambridge, MA: Cambridge University Press.
- Wilkins, D. (1980). Using pattern and plans in chess. *Artificial intelligence*, 14, 165-203.
- Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory and Cognition*, 13, 193-201.

Légende des figures

Figure 1: Illustration d'un réseau de discrimination. Dans certains modèles, tels EPAM, seuls les noeuds terminaux contiennent une image.

Figure 2: Vue d'ensemble d'EPAM/CHREST. Les inputs visuels et acoustiques sont dirigés vers la mémoire à long-terme (MLT) par un réseau de discrimination (RdD) visuel et acoustique, respectivement. L'information obtenue en MLT est utilisée pour effectuer une action motrice, pour être encodée en mémoire à court-terme (MCT) visuelle ou acoustique, ou pour être combinée avec l'input externe et être discriminée à nouveau.

Figure 3: Création de chablon (voir texte).

Figure 4: Création d'une production à partir de deux réseaux de discrimination. Une configuration de pièces sur l'échiquier va suggérer un coup potentiel.

Figure 5: Exemple de la manière dont les arcs latéraux sont créés avec MOSAIC. (Les tests ont été omis dans cette figure)

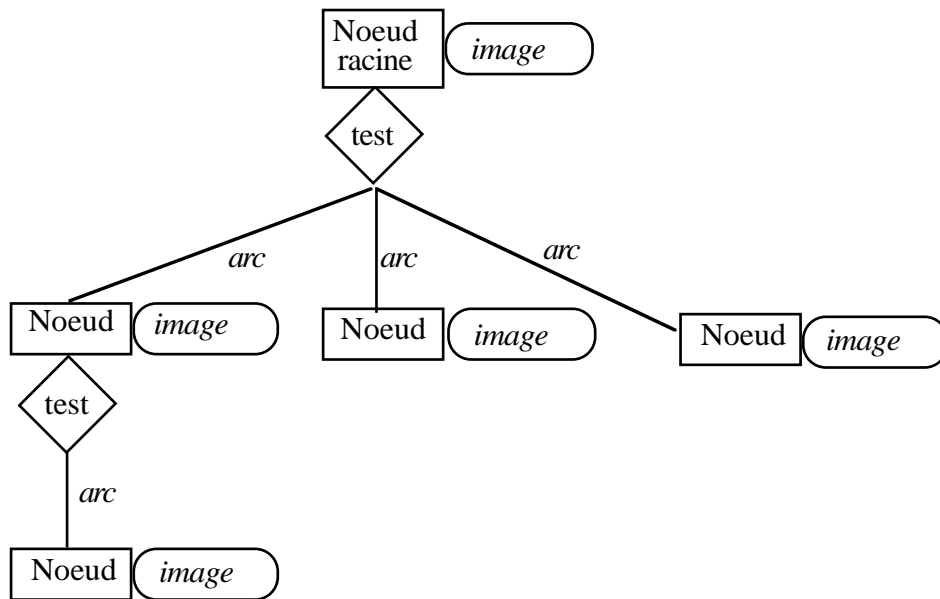


Figure 1

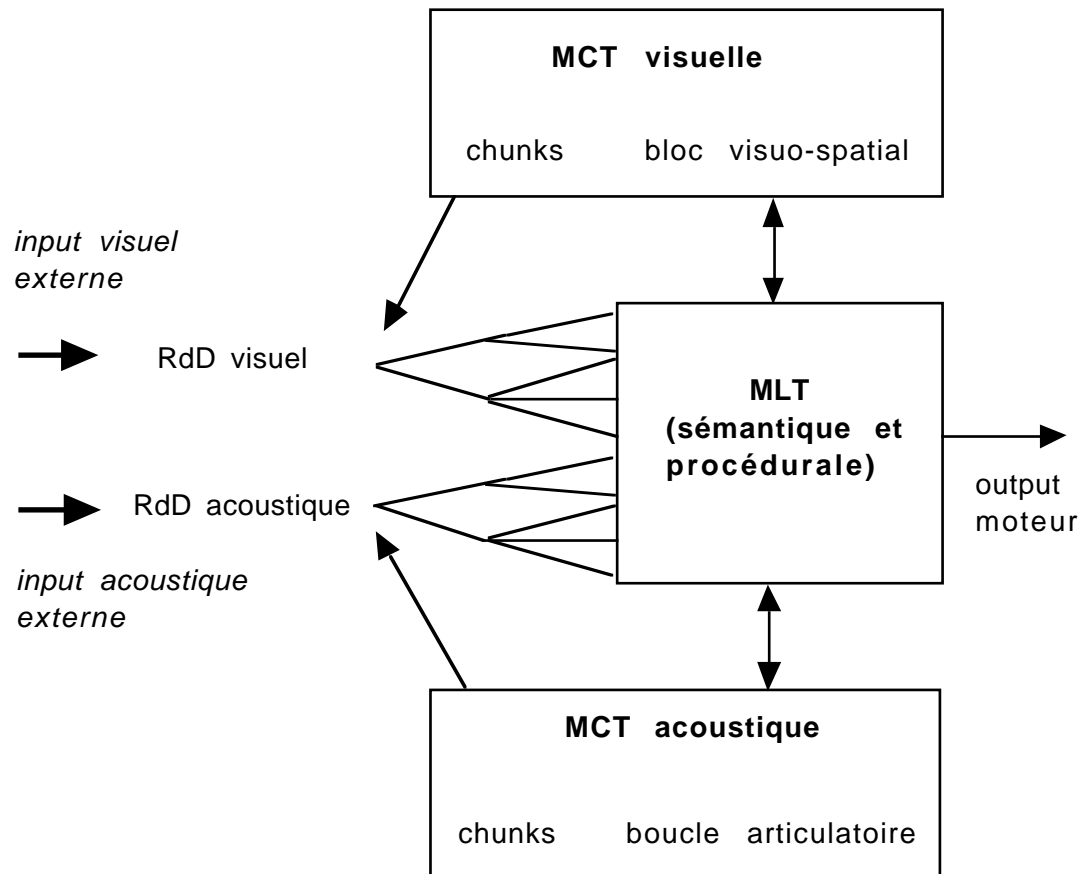


Figure 2

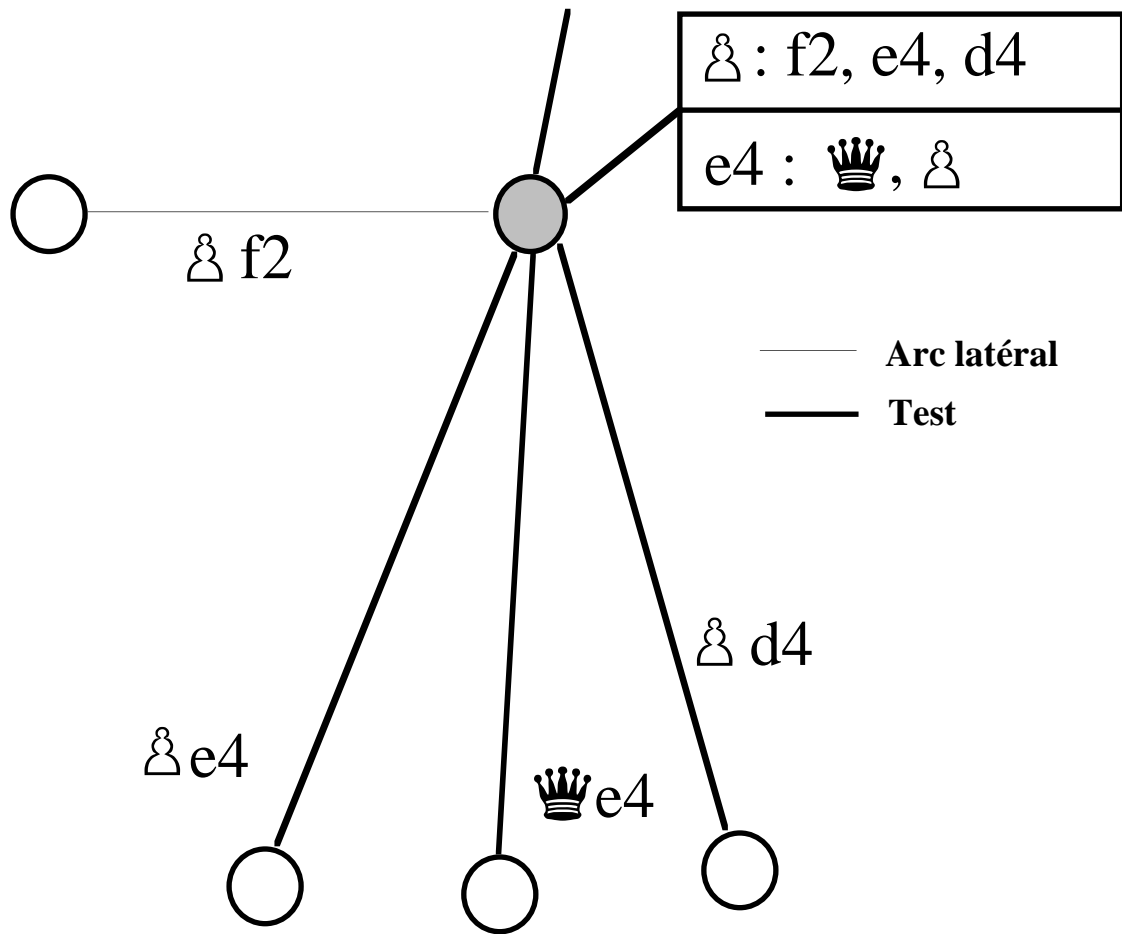


Figure 3

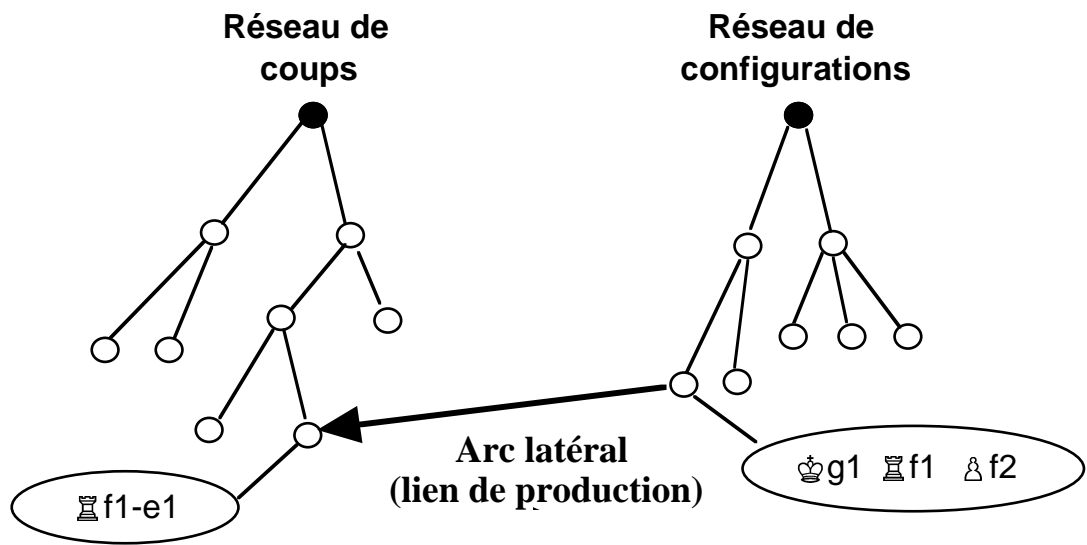


Figure 4

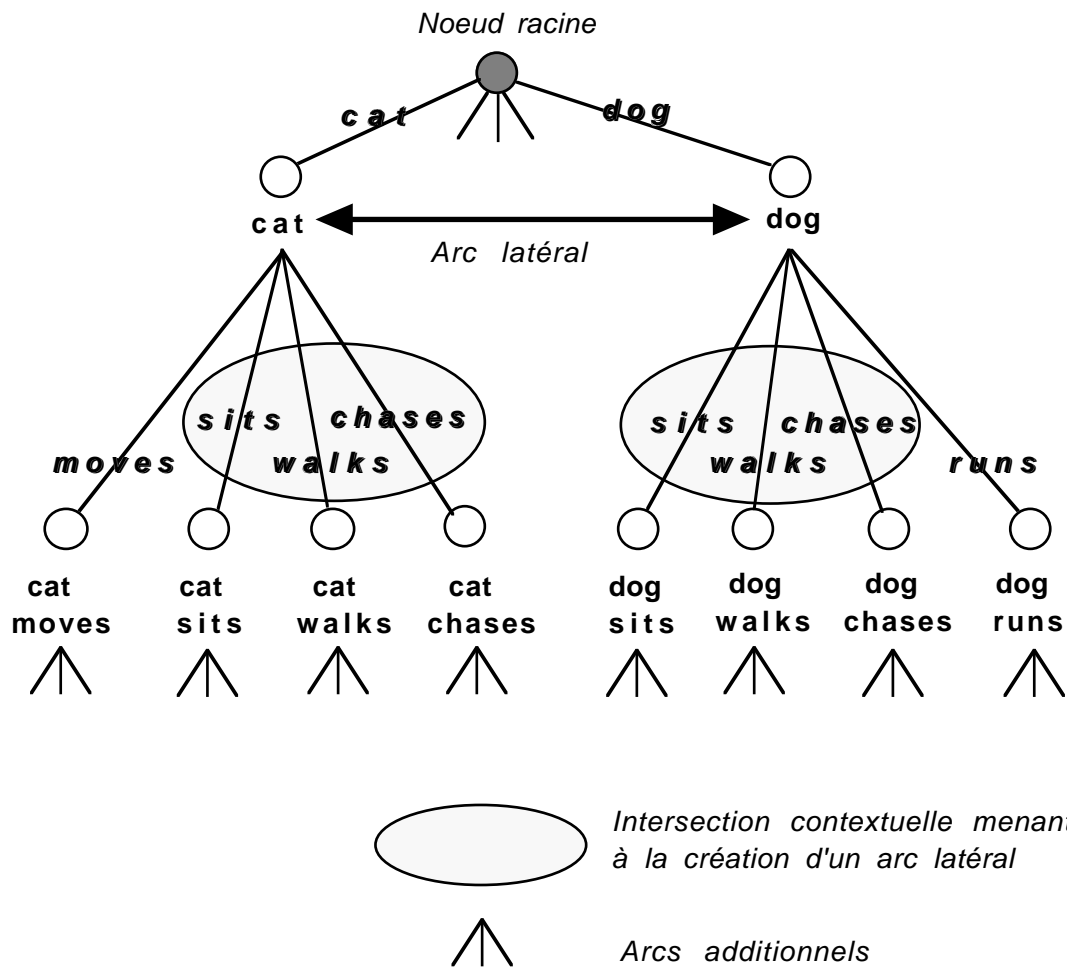


Figure 5

¹CHREST est l'abréviation de Chunk Hierarchy and REtrieval STructures (Hiérarchie de chunks et structures de retrait).

²Ces mécanismes de reconnaissance n'ont pas été implémentés dans les diverses versions d'EPAM, le système étant supposé connaître un ensemble de primitives correspondant aux propriétés élémentaires des objets.

³ Pour une manière alternative de générer des règles de production à partir d'un réseau de discrimination, voir Quinlan (1986).