

Chapitre 8 Mémoire et apprentissage

Pierre Perruchet et Bénédicte Poulin-Charronnat

La littérature scientifique sur la mémoire et l'apprentissage est immense. La synthèse en quatre volumes récemment éditée par John Byrne (2008), qui totalise plus de 3 000 pages, permet de se représenter l'étendue du domaine: chacun de ses 157 chapitres n'est lui-même qu'une synthèse d'un ensemble de travaux souvent vaste. L'objectif du présent volume d'illustrer la complémentarité de différentes approches impose certes une première sélection. Par exemple, l'analyse des mécanismes moléculaires et cellulaires, qui constitue plus du quart de l'ouvrage précité, ne sera pas évoquée ici, en raison de l'insuffisance des points de contact entre cette analyse et l'approche cognitive. Mais il reste une multitude d'approches, de théories et de données expérimentales, au sein desquelles le choix ne peut qu'être arbitraire et subjectif.

C'est donc sans aucune prétention d'exhaustivité que nous traiterons de deux questions. La première concerne ce qu'il est convenu d'appeler les « systèmes » de mémoires. Dans une première grande section, nous poserons le problème et examinerons pourquoi les recherches récentes ont conduit à relativiser l'importance de certaines distinctions classiques, en l'occurrence entre mémoire à court terme et à long terme, d'une part, et entre mémoire explicite et implicite, d'autre part. La section suivante mettra l'accent sur la distinction qui paraît aujourd'hui recueillir le plus large consensus, entre un système hippocampique essentiellement dédié à la mémoire précise d'épisodes spécifiques et un système néocortical configuré pour exploiter les régularités statistiques de l'environnement.

La seconde question traitée est relative aux apprentissages statistiques. Nous examinerons dans les deux dernières sections pourquoi et comment

la prise en compte des régularités statistiques présentes dans l'environnement est devenue centrale pour la psychologie contemporaine de l'apprentissage. Tout au long de ces analyses, différentes approches des sciences cognitives seront considérées. Nous concluons sur quelques propositions concernant leur complémentarité, mais sans doute est-il plus simple pour le lecteur de dévoiler dès à présent l'idée principale que nous voudrions illustrer. En bref, les données de la psychologie expérimentale traditionnelle et les données de la neuropsychologie sont sources d'hypothèses et de modèles, faisant l'objet d'une croissance continue en termes de complexité et de diversité. Face à cela, la modélisation computationnelle, qu'elle soit d'inspiration connexionniste ou non, tend à exercer un rôle modérateur et unificateur, en suggérant comment ces données peuvent être interprétées de façon plus économique. Plus précisément, notre fil conducteur sera d'illustrer comment la modélisation permet de démontrer la puissance ou la viabilité de modes d'explication qui, exprimés sous une forme purement verbale, pourraient paraître insuffisants ou inadaptés pour rendre compte des données empiriques.

Les « systèmes » de mémoires

Pour comprendre le problème que le concept de « systèmes » de mémoires vise à résoudre, il suffit d'imaginer la diversité des situations mettant en jeu des processus d'acquisition. La phase d'étude ou de familiarisation peut être intentionnelle (en laboratoire, le sujet sait que l'on teste sa mémoire) ou incidente (un prétexte est donné pour justifier la tâche, comme juger du caractère plaisant des items présentés). La tâche peut impliquer du matériel visuel, tel que des mots ou des images, ou auditif, ou encore consister à acquérir une habileté motrice. Il peut d'agir d'une liste d'items simples ou d'une liste de paires d'items (une littérature considérable existe sur l'apprentissage de « paires associées », mauvaise traduction de *paired associates*). L'intérêt peut porter sur les items eux-mêmes ou sur leur ordre dans une liste (les mécanismes d'apprentissage sériel font l'objet d'approches spécifiques). La mémoire peut être testée au moyen d'une tâche de rappel libre (par exemple, quel mot avez-vous

vu ?), de rappel dit « indicé » (quel mot avez-vous vu qui commence par « ta » ?) ou de reconnaissance (avez-vous vu « table » ou « tableau » ?). L'intérêt peut porter sur la mémoire d'événements passés, ce qui est le cas le plus fréquent, mais également sur ce que l'on doit faire à l'instant présent : c'est la mémoire prospective, qui a récemment donné lieu à de nombreuses études. La liste des variantes possibles est évidemment très loin d'être exhaustive.

Le problème posé par cette diversité est le suivant. Si l'on observe l'effet d'une variable quelconque (par exemple, le fait d'espacer la répétition des items dans une liste) dans un apprentissage intentionnel avec des mots présentés visuellement, en utilisant un test de reconnaissance présenté 10 minutes après la phase de familiarisation, à quelle autre situation pourra-t-on généraliser ces résultats ? Si l'on observe un déficit dans cette tâche pour une catégorie de la population (par exemple, des personnes âgées), que pourra-t-on conclure sur l'étendue du déficit ? La réponse impose de catégoriser les situations, et il est usuel de qualifier les catégories ainsi obtenues de « systèmes ». Ce que doit être un système de mémoire a fait l'objet de nombreux débats dans lesquels il n'est pas possible d'entrer, mais de façon générale, outre le fait de posséder une organisation particulière et un certain nombre de principes communs, un système de mémoire est le plus souvent associé à une structure anatomique spécifique, ou du moins à un réseau neuronal physiquement identifiable.

Toute la question est évidemment de définir le nombre et la nature des systèmes qu'il convient de retenir. Selon les auteurs, la réponse peut être un seul, ce qui revient à nier l'utilité du concept, plus classiquement deux, avec des oppositions comme court terme/long terme, explicite/implicite, déclaratif/procédural, épisodique/sémantique, et bien d'autres jusqu'à... 256, nombre qu'Endel Tulving, une figure majeure dans ce champ d'étude, a proposé avec un peu d'humour (mais, la liste de 256 est bien présente en annexe ! Cf. Tulving 2007). On trouve également des classifications comprenant de l'ordre de 5 à 10 catégories, avec diverses tentatives d'organisation hiérarchique. Nous en resterons aux classifications binaires, et tout d'abord à deux d'entre elles qui ont eu une importance historique certaine : les oppositions court terme/long terme et explicite/implicite.

Mémoire à court terme et mémoire à long terme

L'idée selon laquelle les processus responsables de la rétention d'information sur des délais de quelques secondes diffèrent de ceux qui régissent la mémoire sur des délais se comptant en minutes, heures, jours, années est très ancienne. Mais l'idée de systèmes spécifiques a surtout pris corps après que Scoville et Milner eurent publié, en 1957, le cas du patient H. M. Ce patient avait subi une ablation bilatérale des lobes temporaux médians afin de traiter une épilepsie sévère. Le traitement contre l'épilepsie se révéla efficace, mais les effets sur la mémoire furent désastreux. Spécifiquement, H. M. avait perdu toute capacité de se former de nouveaux souvenirs. Ses performances dans des tâches de mémoire n'excédant pas quelques secondes de rétention étaient toutefois comparables à celles de sujets normaux. Par la suite, d'autres patients furent identifiés comme présentant la configuration inverse, à savoir une mémoire à court terme (MCT) déficiente liée à une mémoire à long terme (MLT) intacte (Shallice et Warrington 1970). Cette « double dissociation » a peu à peu renforcé l'idée selon laquelle MCT et MLT reposeraient sur deux aires de stockage neurologiquement distinctes. Des modèles psychologiques ont été élaborés, et la distinction entre MCT et MLT a servi d'explication à un certain nombre d'effets, comme, par exemple, les effets liés à l'ordre de présentation des items dans une liste. On sait que de façon générale, les items situés dans le corps de la liste sont moins bien mémorisés que les items situés en début de liste (effet de primauté) et les items situés en fin de liste (effet de récence). Parce que l'effet de récence disparaît en cas de rappel différé (Postman et Phillips 1965), à l'inverse de l'effet de primauté, ce dernier a été relié à la MLT, alors que l'effet de récence a été relié à la MCT.

La démonstration apportée par la neuropsychologie s'est progressivement avérée beaucoup moins claire qu'il ne le semblait. Les raisons en sont multiples, mais la principale est sans doute qu'un certain nombre de facteurs ont été initialement confondus avec la variable « durée de rétention », et notamment le type d'information traitée. Par exemple, les tests visant à mesurer la MCT emploient souvent un matériel plus simple (ou plus familier) que ceux mesurant la MLT (pour une synthèse, cf. Ranganath et Blumenfeld 2005). Ainsi, les lobes temporaux médians

(ceux qui avaient été retirés chez H. M.) semblent nécessaires à la rétention à court terme dès lors qu'un matériel plus complexe est utilisé pour la mesure. Les explications psychologiques qui reposaient sur la dualité MCT/MLT ont, elles aussi, changé. Ainsi, pour les effets d'ordre évoqués plus haut, il est apparu que l'effet de récence pouvait persister à long terme sous certaines conditions expérimentales, rendant caduque l'idée selon laquelle il attesterait d'une MCT autonome (Crowder 1993).

Peut-on alors considérer que la distinction entre MCT et MLT est obsolète? Ce serait aller trop loin, car la distinction garde au moins un intérêt descriptif. Mais l'idée d'une dualité de systèmes, responsables de phénomènes comportementaux spécifiques, n'est plus d'actualité, et la modélisation computationnelle apporte sur ce point un éclairage particulier. Considérons SIMPLE, un acronyme pour *Scale-Independent Memory, Perception, and Learning*. SIMPLE est un modèle proposé par Brown, Neath et Chater (2007), dont l'objectif est de montrer que, moyennant une certaine forme de codage – la trace des événements en mémoire inclut leur distance temporelle au présent – et certains postulats, il devient possible de rendre compte d'un grand nombre de phénomènes au moyen de mécanismes similaires, quelle que soit l'échelle de temps considérée. Par exemple, SIMPLE rend compte aussi bien des cas où l'effet de récence disparaît à long terme que des cas où celui-ci persiste, évitant ainsi de postuler tout à la fois l'existence de deux formes de mémoires, à court et à long terme, et l'existence de deux formes d'effets de récence, chacune étant sous la dépendance de l'un des systèmes. Le même modèle se révèle également capable d'expliquer certaines dissociations neuropsychologiques initialement considérées comme des preuves de dualité MCT/MLT (Brown, Della Sala, Foster et Vousden 2007). Ces résultats ne prouvent pas que MCT et MLT ne constituent pas deux systèmes séparés, mais ils démontrent que les données actuelles ne contraignent pas à faire l'hypothèse d'une telle distinction.

Mémoire explicite et mémoire implicite

À bien des égards, l'idée de systèmes respectivement dévolus à la mémoire explicite et à la mémoire implicite a suivi la même évolution. Là aussi, l'origine est à rechercher dans des dissociations observées dans

les troubles neuropsychologiques, confortées par des dissociations expérimentales chez le sujet sain. La remise en cause s'est là aussi opérée par la découverte d'un certain nombre de facteurs confondus avec la variable initialement tenue pour critique, les modèles computationnels venant confirmer l'idée qu'il n'est nul besoin d'imaginer des systèmes indépendants.

Considérons par exemple l'étude de Keane, Gabrieli, Noland et McNealy (1995), qui exploite une variante des procédures initialement développées par Warrington et Weiskrantz (1968). Des patients amnésiques et des sujets contrôles sont tout d'abord exposés à une liste d'items composés de suites arbitraires de trois consonnes (par exemple, RDP, HRW ou SMN). On leur présente ensuite une autre liste dans laquelle les items d'étude sont mélangés à des items nouveaux, mais similaires. Les sujets doivent dire, pour chaque item, s'il a été vu ou non. Il s'agit d'une tâche de reconnaissance, qui est une tâche standard de mémoire dite « explicite ». Ce qui est explicite est la recherche en mémoire des items présentés que la tâche implique de la part du sujet. Dans une seconde épreuve, les mêmes sujets sont exposés à de nouvelles suites pour étude, mais la tâche de test qui suit diffère. À nouveau des items déjà vus et des items nouveaux sont présentés, mais cette fois de façon très brève, et la tâche des sujets est simplement de lire les items. On observe généralement une différence de performance entre les items vus et non vus dans ce type de tâche, les items vus étant plus facilement identifiés. Cet effet, appelé « amorçage de répétition », est une manifestation de la mémoire – il faut bien qu'une trace de l'expérience antérieure subsiste pour expliquer l'amélioration observée –, mais il s'agit d'une manifestation « implicite » (ou indirecte), puisque l'évocation explicite du souvenir n'est pas requise. Les résultats de l'expérience de Keane *et alii* sont que, de façon non surprenante, les amnésiques ont de moins bonnes performances que les normaux dans la tâche explicite de reconnaissance, mais se comportent normalement dans la tâche implicite d'amorçage. Ces résultats, qui répliquent nombre d'études antérieures, ont été considérés comme une preuve de l'existence de deux systèmes neurologiquement distincts, respectivement consacrés aux mémoires explicite et implicite. Nombre de dissociations comportementales ont également été mises en évidence chez les sujets normaux.

Des simulations ultérieures ont toutefois montré que la preuve ainsi faite était loin d'être convaincante. Ainsi, Kinder et Shanks (2003) ont tenté de simuler les résultats de Keane *et alii* (1995) au moyen d'un simple réseau connexionniste. La seule variation qu'ils introduisent pour reproduire la différence entre patients amnésiques et sujets contrôles porte sur le taux d'apprentissage : un taux d'apprentissage faible est censé correspondre au traitement des amnésiques et un taux d'apprentissage élevé à celui des contrôles. Ce que montrent les auteurs est que cette simple variation, à l'exclusion de tout autre changement de structure ou de paramètre, suffit à expliquer la dissociation observée par Keane *et alii*. Quand les items tests sont présentés au réseau dans le même format que les items d'étude, comme dans la tâche de reconnaissance, l'effet du taux d'apprentissage est important. En revanche, quand les items tests sont présentés de manière parcellaire ou très rapide, comme dans la tâche d'amorçage, l'effet du taux d'apprentissage devient nul ou négligeable. Pour les auteurs, ce phénomène reflète une propriété générale des modèles connexionnistes, à savoir que l'impact du taux d'apprentissage est beaucoup moins important lorsque les items d'étude et de test diffèrent que lorsque les items d'étude et de test sont identiques. Ce que fournit l'étude de Kinder et Shanks est une démonstration de principe que les dissociations neuropsychologiques entre mémoires explicite et implicite ne peuvent en aucun cas démontrer l'existence de systèmes neurologiques distincts : de telles dissociations peuvent être reproduites en ne faisant varier qu'un paramètre dans un réseau unique (pour une brève synthèse sur l'intérêt de la modélisation pour démontrer la possibilité de modèles unitaires des mémoires explicite et implicite, cf. Berry, Shanks et Henson 2008).

On a pu également montrer que le caractère explicite ou implicite des épreuves était confondu avec un certain nombre de caractéristiques. Ainsi, la plupart des épreuves implicites recourent à des tâches essentiellement perceptives, dites aussi *bottom-up*. Il s'agit en effet très souvent de montrer que l'exposition antérieure à un stimulus facilite la perception ultérieure de ce même stimulus (ce qui se traduit par une réduction de la latence ou du nombre d'erreurs) lorsqu'il est présenté dans des conditions dégradées. La plupart des épreuves explicites recourent au contraire à des tâches appelées, par contraste, « conceptuelles » ou

top-down, puisque l'essentiel de la tâche consiste à rechercher l'information en mémoire. Les distinctions sont beaucoup moins nettes dès lors que l'on tente de contrôler ces facteurs (Blaxton 1989). Il ne s'agit certes pas de conclure que des tâches explicites et implicites de mémoire peuvent être utilisées indifféremment, mais l'idée que les différences observées entre ces tâches attestent de l'existence de systèmes distincts a sans doute moins de défenseurs aujourd'hui que par le passé.

Vers deux systèmes complémentaires : données comportementales, neuropsychologiques et computationnelles

On peut comprendre que les distinctions envisagées jusqu'à présent ne renvoient pas à l'existence de systèmes de mémoire autonomes. En effet, si l'on raisonne en termes de nécessité adaptative, il ne semble pas y avoir d'incompatibilité fondamentale entre les fonctions servies par ces systèmes hypothétiques. Par exemple, il est difficile de voir pourquoi il serait nécessaire d'avoir développé des systèmes séparés pour le court et le long terme, sachant que l'objectif final est le même : récupérer l'information initiale aussi fidèlement que possible. Mais quand l'exploitation des expériences antérieures à des fins adaptatives exige-t-elle réellement le développement de plusieurs systèmes de mémoire ? Une réponse possible est la suivante : lorsque l'on vit une expérience donnée, que l'on appelle communément un « épisode » dans ce contexte, il est utile d'en retirer au moins deux types d'information. D'une part, il est généralement utile d'en garder un souvenir spécifique, daté et contextualisé, et, d'autre part, il est également souhaitable que cet épisode enrichisse notre représentation générale du monde. Si je suis attaqué par un chien errant, il pourra m'être utile de conserver un souvenir précis de l'événement, ne serait-ce que pour pouvoir en faire état à qui de droit, mais il paraît également utile que mon interaction avec ce nouvel animal participe à l'élaboration de ma connaissance générale des chiens, notamment par la mise à jour de ma représentation des liaisons statistiques existant entre des traits physiques (par exemple, les oreilles dressées) et des traits comportementaux (la dangerosité). Le point crucial est que ces

deux fonctions semblent *a priori* incompatibles, car la première requiert une conservation des détails et du contexte, alors que la seconde implique que les informations spécifiques individualisant l'épisode soient négligées au profit de l'extraction de régularités statistiques.

Est-il réellement impossible d'optimiser les deux fonctions au sein d'un seul système ? Certains modèles ont cette prétention. Dans MINERVA 2 (Hintzman 1986), chaque épisode individuel forme une trace indépendante, quelle que soit sa similarité avec un épisode antérieur. Même strictement identiques, deux événements font l'objet de deux traces séparées. Le « magasin de souvenirs » ainsi engendré est supposé être accessible en parallèle à tout moment, soit pour retrouver une trace individuelle, soit pour abstraire une information plus générale, d'ordre catégoriel, par exemple. Le modèle permet ainsi de simuler adéquatement de nombreux résultats expérimentaux. On peut toutefois douter de son intérêt réel. Si l'on sort du contexte limité du laboratoire auquel le modèle a été appliqué, MINERVA 2 apparaît requérir une capacité de stockage extraordinaire, ainsi qu'une puissance de traitement lors de la récupération tout aussi démesurée. Il ne paraît pas trop surprenant qu'un modèle qui ne pose aucune limite dans la capacité de stockage et la puissance de traitement auxquelles il accède parvienne à rendre compte d'un grand nombre de données (voire de toutes les données), mais son pouvoir explicatif vis-à-vis des processus psychologiques réellement mis en jeu chez l'homme semble limité.

Les modèles connexionnistes de la mémoire, tels que le modèle évoqué plus haut (Kinder et Shanks 2003), apparaissent considérablement plus économiques qu'un modèle comme MINERVA 2. Mais le point crucial pour notre propos est que les réseaux qui réussissent parfaitement dans leur tâche d'extraction et de généralisation ont des performances désastreuses lorsqu'il s'agit de récupérer une information précise et fidèle, et inversement. Il existe une grande variété d'architectures de réseaux qu'il ne s'agit pas d'examiner ici, mais dès lors que l'objectif est de simuler le comportement humain, les réseaux les plus utilisés sont les réseaux distribués, du moins au niveau de la couche intermédiaire (le codage des entrées et sorties pouvant être localiste). En effet, ces réseaux rendent compte naturellement d'un grand nombre de capacités humaines que les modèles symboliques peinent à reproduire. Ainsi,

ils peuvent compléter un stimulus dont seule une partie incomplète ou dégradée est présentée; ils sont également capables de répondre de façon identique à des stimulations proches, mais différentes, manifestant ainsi des capacités de catégorisation et de généralisation. En bref, ils se révèlent sensibles, tout comme les humains, à l'information statistique présente dans l'environnement. Malheureusement, ces réseaux sont bien peu performants lorsqu'il s'agit de retrouver une information précise et contextualisée. En effet, le taux d'apprentissage, qui détermine l'importance des modifications induites dans le réseau à chaque nouvel événement, a besoin d'être faible pour que le réseau se configure en fonction de l'ensemble des données du domaine considéré (et non en fonction des seuls derniers événements stables), ce qui se révèle non adapté à la formation du souvenir d'un événement unique. De plus, ces réseaux sont sensibles à ce qu'on appelle l'« interférence catastrophique »: de par le codage distribué et le mode d'apprentissage, les modifications induites par un stimulus donné se trouvent irrémédiablement altérées par la présentation d'autres stimuli. L'interférence est également une marque de la mémoire humaine, mais dans une proportion considérablement moindre. Si les informations potentiellement interférentes sont présentées de façon mixée et répétée, le réseau peut apprendre à les différencier, mais cette solution n'est pas applicable aux souvenirs épisodiques, dont la propriété essentielle est, précisément, d'être uniques.

Il est possible d'optimiser le codage d'informations épisodiques, mais d'autres structures de réseaux sont alors nécessaires, qui maximisent l'individuation des événements et leur rapidité d'encodage. En bref, il n'apparaît pas possible d'optimiser simultanément la découverte de la structure globale d'un ensemble d'événements et d'expériences et la formation rapide du souvenir précis d'un épisode particulier. Il s'agit là d'objectifs fonctionnellement incompatibles, dans le sens que Sherry et Schacter (1987) donnent aux processus d'adaptation qui requièrent l'implication de plusieurs systèmes.

Des distinctions relativement similaires à celles-ci ont été proposées depuis longtemps, contrastant en particulier mémoire déclarative et mémoire procédurale (Cohen et Squire 1980) ou mémoire épisodique et mémoire sémantique (Tulving 1972). On peut débattre à l'infini

du choix des termes, et il est inévitable qu'une terminologie particulière, en mettant l'accent sur une dimension donnée, ait un caractère réducteur. À la suite de Sherry et Schacter (1987), nous désignerons ici par Système I le système impliqué dans l'acquisition d'habiletés perceptives, cognitives et motrices, de connaissances sémantiques et catégorielles, c'est-à-dire, de façon générale, regroupant tous les processus qui procèdent par ajustement graduel aux propriétés statistiques de l'environnement, et par Système II le système responsable de l'apprentissage d'épisodes uniques et contextualisés, susceptibles d'être explicitement réévoqués dans leur singularité. Un très grand nombre de données, provenant tant de la neuropsychologie que de l'imagerie cérébrale, permettent de rattacher ces systèmes à des aires cérébrales différentes. Brièvement, le Système II est classiquement rattaché aux aires hippocampiques, alors que le Système I semble lié au néocortex dans son ensemble et à certaines structures sous-corticales telles que les ganglions de la base.

Si cette distinction, quelle que soit la terminologie retenue, est celle qui reçoit actuellement le plus large consensus, en raison de la remarquable convergence des informations issues des différentes approches, il existe également un consensus pour ajouter que ces systèmes ne peuvent être conçus en termes de modules indépendants. Ainsi, dans la théorie d'Anderson (1983), il y aurait un passage graduel du « déclaratif » au « procédural » lors de l'acquisition des habiletés. L'idée selon laquelle la mémoire sémantique serait formée par le résidu des traces épisodiques a aussi été défendue. Ces théories suggèrent que la formation du Système I (intégrant les notions de procédural et de sémantique) dépend de l'activité du Système II (proche du versant déclaratif et épisodique). Mais les interactions pourraient être plus complexes que les relations unilatérales décrites dans ces travaux anciens. L'idée défendue par les tenants des « systèmes d'apprentissage complémentaires » est, certes, que l'information est bien transférée de l'hippocampe au système cortical, le premier jouant le rôle d'un instructeur ou d'un « répéteur » pour le second (McClelland, McNaughton et O'Reilly 1995; O'Reilly et Rudy 2001). Mais l'information transférée n'est pas une copie brute de l'information sensorielle: une propriété cruciale de l'organisation du système cortical est que les entrées dans les lobes temporaux médians

sont elles-mêmes en provenance des aires néocorticales. Le schéma qui s'esquisse est donc que la formation des souvenirs événementiels dans le Système II – souvenirs qui participeront ultérieurement à l'ajustement du Système I aux régularités statistiques de l'environnement – dépend elle-même des représentations condensées, abstraites, élaborées antérieurement par le néocortex. Les informations en provenance de différentes régions corticales par les projections cortico-hippocampiques, nécessairement multifformes puisqu'elles sont formées des éléments hétérogènes composant une expérience singulière, seraient intégrées en une représentation cohérente (le processus de *binding*), avant de retourner enrichir le Système I de cette nouvelle configuration d'événements.

Plus profondément encore, cette architecture interactive conduit à réviser l'idée même de systèmes « de mémoires ». L'image de la mémoire comme une bibliothèque de souvenirs, alimentée en permanence par de nouvelles perceptions provisoirement stockées en MCT, ne peut plus guère être maintenue. Comme l'exprime Lynn Nadel (2008), l'idée d'un système cérébral uniquement dévolu à la MLT n'a plus de sens : l'ensemble des systèmes neuronaux sont sans doute impliqués tout à la fois dans le traitement et la mémorisation. En d'autres termes, ce qui est perçu et mémorisé à un instant donné dépend étroitement des expériences antérieures et vient à son tour modifier la façon dont les épisodes futurs seront eux-mêmes perçus et mémorisés.

Les apprentissages statistiques

Dans la section précédente, nous avons qualifié de « statistiques » les apprentissages réalisés par le Système I, résumant par ce terme l'ensemble des acquisitions qui reposent sur l'extraction d'invariants d'expériences multiples. L'usage du terme en ce sens est relativement récent, mais il rencontre un succès grandissant, sans doute en raison d'une double évolution. D'une part, il est apparu dès le tournant des années 1970 que les apprentissages les plus élémentaires, tels que le conditionnement pavlovien, étaient en fait plus complexes qu'on ne le pensait. En particulier, les recherches ont montré que la simple contiguïté temporelle ou spatiale de deux événements n'était pas suffisante pour

qu'un conditionnement s'élabore. Il est nécessaire que l'occurrence de l'un des deux événements puisse être effectivement prédite à partir de l'autre, ce qui suppose l'existence d'une relation de contingence entre les événements, telle qu'elle peut être exprimée par un coefficient de contingence standard comme le Chiz. D'autre part, et c'est le point que nous tenterons d'illustrer dans cette section, il est également apparu que des apprentissages très élaborés, semblant impliquer la découverte de règles complexes, pouvaient en fait relever d'une sensibilité à des régularités statistiques relativement élémentaires. Cela concerne virtuellement toutes les formes d'apprentissage, mais le domaine central d'intérêt pour de nombreux chercheurs est l'acquisition du langage naturel, que la tradition chomskyenne invite à penser comme hors de portée de tout mécanisme d'apprentissage, quel qu'il soit, parce qu'elle reposerait sur l'acquisition de règles que les données dont l'apprenant dispose ne permettent pas d'abstraire.

Nous prendrons ici pour illustration le débat qui dure depuis plus de vingt ans, concernant l'apprentissage du passé en anglais. Comme chaque francophone l'a appris à l'école, la formation du temps passé répond à une règle, que l'on peut formuler comme « ajouter «-ed» à la racine des verbes réguliers ». L'idée traditionnelle est que c'est exactement cette règle que le petit anglophone applique lorsqu'il apprend à parler, ce que vient apparemment confirmer le phénomène de surgénéralisation (c'est-à-dire le fait que durant un certain temps, l'enfant tend à appliquer la règle du «-ed» à des verbes irréguliers, engendrant par exemple «goed» au lieu de «went»). En effet, ce phénomène atteste que le suffixe «-ed» n'est pas lié à un item spécifique, mais à une variable, à savoir la classe syntaxique des verbes réguliers (la régularisation étant conçue comme l'inclusion erronée de certains verbes irréguliers au sein de cette variable).

Le défi le plus sérieux à cette vision traditionnelle est, à nouveau, venu de la modélisation connexionniste. En réponse à l'affirmation selon laquelle l'acquisition du langage requiert l'acquisition de règles, l'approche connexionniste avance que des modèles n'incluant aucune règle, et seulement composés d'un réseau d'associations plus ou moins fortes, sont capables d'apprendre des aspects fondamentaux du langage. Pour l'apprentissage du temps passé, Rumelhart et McClelland (1986)

ont montré qu'un réseau connexionniste était capable d'apprendre le temps passé des formes régulières et irrégulières à partir d'une représentation phonologique des verbes, et qu'il parvenait même à reproduire l'apparition transitoire de la surgénéralisation. Cette étude initiale a été fortement critiquée par les défenseurs de la position classique. Cette critique a conduit à une évolution des modèles et, depuis, ces échanges se sont poursuivis sans relâche, les modélisateurs ajustant leurs modèles pour répondre aux défis successifs posés par les défenseurs d'une vision traditionnelle (McClelland et Patterson 2002). Des débats similaires, bien que plus limités, ont eu lieu sur d'autres aspects du langage. Il n'est pas possible d'entrer ici dans le détail des arguments. Nous voudrions plutôt tenter de dégager les aspects qui permettent de comprendre l'essence et la puissance de l'approche statistique. En effet, il ne faudrait pas comprendre « statistique » dans le sens d'« appauvri », donnant de l'apprentissage statistique l'image d'un apprentissage bruité ou limité, par exemple, aux cas les plus faciles ou les plus fréquents. Deux arguments principaux s'opposent à cette vision.

Le premier point est que, dans l'approche statistique, l'individu n'apprend pas une « version probabiliste des règles » ; il apprend généralement quelque chose qui se trouve être à un niveau de description autre que ce que les règles décrivent. Ce qu'il apprend, ce sont des éléments facilement perceptibles qui se trouvent, pour des raisons parfois très indirectes, corrélés aux règles. Si un réseau connexionniste est capable d'apprendre le temps passé pour les verbes réguliers et irréguliers, c'est parce que les verbes réguliers et irréguliers diffèrent dans la distribution de leurs traits phonologiques : par exemple, la majorité des verbes phonologiquement similaires à « drink »/« drank » ont également une flexion irrégulière (« sing »/« sang », « ring »/« rang », « stink »/« stank »). Le fait que l'apprentissage porte généralement sur une autre dimension que celle régie par les règles a été amplement attesté dans un champ de recherche qui n'est pas particulièrement orienté vers le langage, mais qui est néanmoins pertinent pour notre sujet, à savoir le champ de l'« apprentissage implicite ». Dans les études expérimentales sur l'apprentissage implicite, des volontaires sont d'abord exposés à du matériel structuré par des règles arbitraires complexes, puis soumis à des tests destinés à évaluer dans quelle mesure ils sont devenus sensibles à cette structure.

Le résultat général est qu'une amélioration des performances est manifeste, alors même que les sujets restent incapables de verbaliser les règles du domaine. L'histoire récurrente dans ce champ de recherche est que les auteurs des premières recherches explorant une situation particulière considèrent, immanquablement, que les sujets ont acquis (inconsciemment) les règles structurant la situation. Ces publications sont régulièrement suivies d'autres publications montrant que les performances ne sont pas dues à l'apprentissage des règles elles-mêmes, mais à la considération d'éléments de surface facilement perceptibles, dont les propriétés statistiques se trouvent être un produit dérivé des règles, pour des raisons parfois difficilement identifiables (pour une synthèse, cf. Perruchet 2008).

Le second point important est que les traits corrélés peuvent être nombreux, et situés à différents niveaux de description. Si l'on considère à nouveau l'exemple du passé en anglais, il est apparu rapidement que les traits phonologiques qui étaient considérés exclusivement par Rumelhart et McClelland (1986) n'étaient pas suffisants pour tout expliquer. Ainsi, comment expliquer qu'il soit possible de former le passé correct de verbes homophones (comme « brake » et « break ») dont l'un est régulier (« brake » → « braked ») et l'autre irrégulier (« break » → « broke ») ? Ramscar (2002) a montré que ceci était possible à condition de considérer des indices additionnels, d'ordre sémantique. Des simulations connexionnistes ont attesté que l'exploitation d'indices multiples, tant phonologiques que sémantiques, conduit à une excellente approximation de la performance qui résulterait de la connaissance formelle de la règle du suffixe « -ed », jointe à la connaissance des exceptions, et permet dans certains cas un meilleur ajustement aux données (Woollams, Joanisse et Patterson 2009). Pour ne considérer qu'un autre exemple de la même approche, il a été montré que la considération de simples statistiques de co-occurrence entre mots conjoints (Redington, Chater et Finch 1998) ou encore d'indices phonologiques (Monaghan, Chater et Christiansen 2005) est riche d'informations sur les catégories grammaticales. Ces recherches, ainsi qu'un large éventail d'autres études (pour une analyse portant sur le langage écrit, cf. Pacton, Fayol et Perruchet 2005), suggèrent que, pour autant qu'il s'agisse de langage, les catégories abstraites sont souvent associées à une configuration statistique particulière de traits perceptivement saillants situés à des niveaux de description variés.

Le développement de cette perspective repose bien sûr sur l'avancée des recherches concernant les processus d'apprentissage humain, mais aussi sur le développement d'un mode de description non conventionnel de l'objet d'étude. En ce qui concerne le langage, cette démarche suppose de transcender les frontières entre domaines que la linguistique traditionnelle considère comme séparés, tels que prosodie, phonologie, syntaxe et sémantique. Si les analyses à venir sur des corpus de langage confirment et étendent les conclusions actuelles, et si le même type d'analyse se révèle heuristique dans d'autres domaines de compétences, alors cette approche devrait s'affirmer comme la première alternative réellement viable à la vision innéiste qui caractérise encore tout un pan de l'approche cognitive. Mais il est trop tôt pour tirer des conclusions définitives, et la seule affirmation possible aujourd'hui est que le champ de pertinence des apprentissages statistiques s'est étendu à des domaines qui semblaient hors de sa portée il y a vingt ou trente ans.

Apprend-on comme les réseaux connexionnistes ?

La conclusion selon laquelle la pertinence d'une approche connexionniste s'étend à des champs de recherche de plus en plus vastes suggère naturellement que les réseaux connexionnistes fournissent un modèle adéquat du fonctionnement de notre cerveau. En d'autres termes, le succès de ces modèles semble attester de leur validité biologique et psychologique en tant que modèles de fonctionnement pour une large part de nos apprentissages. L'idée que nous voudrions présenter pour terminer est que si une telle conclusion est possible, et même sans doute communément acceptée dans la communauté scientifique, elle n'est pas la seule. Une façon alternative d'interpréter les données part du fait que le succès des modélisations connexionnistes vient de la capacité des réseaux à extraire des statistiques. Un tel succès indique que l'adaptation à une situation particulière peut s'opérer par la seule exploitation des régularités statistiques présentes dans cette situation. Ce point est peu contestable et constitue une avancée remarquable. Toutefois, formellement, ce succès n'indique rien sur la façon dont s'opère l'exploitation des régularités statistiques chez un sujet humain. Certes, l'analogie

apparente entre un réseau artificiel et un réseau neuronal peut conduire à penser que l'un fournit le modèle adéquat de l'autre, mais d'autres possibilités existent, que nous introduirons à travers un exemple. Nous verrons ensuite les raisons de préférer l'un ou l'autre modèle.

L'exemple choisi est celui de la découverte des mots du langage. L'une des premières tâches à laquelle les nourrissons doivent faire face pour parvenir à comprendre et émettre du langage consiste à isoler les mots d'un flux plus ou moins continu de parole. Les bébés sont bien sûr aidés dans cette tâche par la présence de différents indices, comme les pauses ou le contour mélodique (Jusczyk 1997). Mais cette information, si elle est utile, ne semble pas nécessaire. En effet, les travaux de Saffran et de ses collaborateurs (Saffran, Aslin et Newport 1996; Saffran, Newport et Aslin 1996) ont clairement démontré que le bébé, dès 8 mois, sait distinguer des « mots » de « non-mots » après une exposition relativement brève à un langage artificiel émis en flux continu, sans pauses ni indices prosodiques. Dans ces expériences, les bébés entendent un langage artificiel formé de quatre « mots » trisyllabiques (par exemple, « golabu » ou « tupiro ») produits en succession immédiate par un synthétiseur de parole, dans un ordre aléatoire (par exemple, « golabutupiropadotibidakutupirogolabu... »). Dans un deuxième temps, ils sont confrontés à des « mots » de ce langage, ainsi qu'à des « non-mots » constitués par de nouveaux assemblages de trois syllabes extraites de la fin d'un mot et du début d'un autre mot (par exemple, « labutu » ou « pirogo »). Les bébés parviennent à différencier les deux catégories d'items au-delà de ce que le hasard prédirait. Des performances semblables ont également été rapportées chez l'enfant et chez l'adulte (Saffran, Newport, Aslin, Tunick et Barrueco 1997). Notons que ce résultat est également généralisable à du matériel visuel (Fiser et Aslin 2001, 2005; Turk-Browne, Jungé et Scholl 2005), suggérant que les processus en jeu pour extraire les mots du langage et les objets de l'environnement spatial pourraient être en partie communs.

Ces performances peuvent s'expliquer aisément en termes d'apprentissage statistique. En effet, dans le langage artificiel utilisé, les probabilités transitionnelles entre deux syllabes contiguës sont plus élevées intra-mot (par exemple, de « go » à « la ») qu'inter-mots (de « bu » à « tu »). Notons que cette propriété reste vraie dans le langage naturel (par exemple,

dans « le tapis », la probabilité que « ta » suive « le » est sans doute plus faible que celle que « pis » suive « ta »), mais seulement en moyenne. De façon non surprenante, les réseaux connexionnistes se révèlent adaptés, et en particulier les SRN (*Simple Recurrent Networks*; Elman 1990). Un SRN est un réseau à couche cachée, qui apprend par rétropropagation de l'erreur. Le langage est entré syllabe par syllabe, et la tâche du réseau est de prédire la syllabe suivante à chaque essai. L'aspect récursif d'un SRN (les poids de la couche cachée à l'essai n sont copiés en entrée lors de l'essai $n + 1$) lui permet de tenir compte dans ses prédictions de plusieurs syllabes antérieures à chaque essai. Il apparaît qu'après entraînement, les prédictions du SRN sont meilleures lorsqu'il s'agit de prédire la syllabe suivante à l'intérieur d'un mot que lorsqu'il s'agit de prédire la syllabe suivant la fin d'un mot. Il est donc possible d'en inférer les frontières de mots, qui sont situées là où l'erreur est maximale (Christiansen, Allen et Seidenberg 1998). Ce qui permet au réseau d'apprendre ainsi est que ses prédictions fournissent une (excellente) approximation des probabilités transitionnelles entre syllabes successives.

Mais le succès des SRN n'implique pas qu'ils constituent la seule classe de modèles possible. Au lieu d'inférer les mots à partir des probabilités transitionnelles entre les syllabes du discours, une autre classe de modèles tente de les découvrir directement, en sélectionnant parmi des unités candidates, appelées *chunks* (d'où le nom de *chunk-based models*). Dans le modèle INCDROP de Brent (1996; Brent et Cartwright 1996), le choix entre les différentes segmentations possibles est traité comme un problème d'optimisation. Le principe de la méthode revient à établir une liste exhaustive de toutes les segmentations possibles (bien que des outils mathématiques évitent au programme de procéder réellement ainsi) et à retenir celle qui permet de satisfaire au mieux trois critères : le nombre de nouveaux mots est réduit au minimum, la somme de leurs longueurs est également réduite au minimum, et le produit de la fréquence relative de tous les mots est le plus grand possible. Ce processus d'optimisation est réalisé par une méthode standard d'inférence statistique fondée sur le principe de « longueur de description minimale » (*Minimum Description Length*, ou MDL). D'autres modèles de segmentation s'appuient également sur cet algorithme (comme le *MDL-Chunker* de Robinet et Lemaire [2009]), parfois dans un cadre bayésien

(Goldwater, Griffiths et Johnson 2009). Ce processus exploite l'apparition progressive des mots dans le lexique, car le calcul d'optimalité tient systématiquement compte de ces derniers.

On peut toutefois douter de la plausibilité psychologique et biologique de tels modèles, qui impliquent qu'un calcul mathématique d'optimalité sur une très large base de données soit réalisé de façon rapide et inconsciente. Le modèle PARSEr proposé par Perruchet et Vinter (1998) recourt à la même stratégie générale que les modèles MDL, à savoir effectuer une sélection parmi une liste d'unités candidates, mais au moyen de processus beaucoup plus simples. D'une part, au lieu d'établir la liste exhaustive de toutes les segmentations possibles d'un corpus donné, les différentes possibilités sont explorées en succession lors de la répétition des mêmes séquences au cours du temps, en se fondant sur le hasard comme source de variabilité. D'autre part, au lieu de sélectionner le mode de segmentation qui engendre le nombre minimum de différents mots, PARSEr exploite un corollaire logique : pour un corpus fixé, minimiser le nombre de mots différents revient à maximiser la fréquence de ces mots. Cet objectif peut être atteint naturellement grâce aux processus d'oubli au cours du temps et d'interférence entre éléments partageant des parties communes. En effet, ces processus concourent à éliminer les unités candidates les moins fréquentes et/ou les moins cohésives. Les unités restantes servent à guider la perception de nouvelles séquences, ce qui assure la convergence du processus et évite la création sans fin de nouvelles unités. Certes, cet algorithme est moins puissant que les méthodes MDL, mais ce point n'est pas crucial dans la mesure où PARSEr semble encore découvrir les mots d'un langage artificiel plus rapidement que les sujets humains (Perruchet et Vinter 1998).

Ainsi, à un niveau général, un même comportement – la découverte des mots d'un flux continu de discours – peut être simulé par des modèles d'architectures aussi différentes que celle d'un SRN, des méthodes MDL et de PARSEr. Comment décider du meilleur modèle ? Le premier critère est évidemment celui de l'adéquation du modèle aux données. Pour ce qui est de la découverte des mots dans la situation originale de Saffran, Newport et Aslin (1996), les différents modèles s'acquittent facilement de cette tâche, ce qui n'est pas spécialement étonnant compte

tenu de la simplicité de la situation. Il faut donc recourir à des situations tests plus spécifiques, une démarche que suivent actuellement un certain nombre d'auteurs (Frank, Goldwater, Griffiths et Tenenbaum 2010; Giroux et Rey 2009; Perruchet et Tillmann 2010; Robinet et Lemaire 2009).

Un autre critère de choix est fondé sur la parcimonie des différentes solutions. Mais il n'existe pas de consensus à ce niveau, car une solution parcimonieuse pour certains ne le sera pas pour d'autres. Même si l'on se limite à une dimension particulière, le choix n'est pas toujours évident. Qu'en est-il par exemple de la plausibilité biologique et psychologique des processus impliqués dans les modèles? De ce point de vue, les modèles basés sur MDL n'ont sans doute pas l'avantage. Mais la comparaison entre un SRN et PARSEr est plus délicate. Certes, un SRN a l'avantage, comme tous les modèles connexionnistes, d'avoir une certaine analogie avec un vrai réseau neuronal, composé de neurones et de synapses. De plus, les réseaux implémentent naturellement l'idée séduisante d'une mémoire reconstructive. Plutôt que d'imaginer la mémoire comme un magasin ou une bibliothèque où l'information, stockée telle quelle, devrait être retrouvée, la mémoire est conçue comme une capacité à engendrer les souvenirs. Ce qui est conservé au cours du temps, ce ne sont pas les souvenirs eux-mêmes, mais la capacité de les recréer dans un certain contexte (en d'autres termes, la capacité de produire une certaine sortie, étant donné une certaine entrée). Les modèles connexionnistes apparaissent de ce point de vue plus satisfaisants qu'un modèle comme PARSEr, dans lequel les *chunks* sont conservés sous forme littérale. Mais d'un autre point de vue, des réseaux tels que les SRN ne se conforment pas à l'organisation interactive des processus de mémoire et d'apprentissage esquissée dans la section «Les "systèmes" de mémoires», p. 252. Si les entrées du Système II, traitant chaque épisode individuel, proviennent des structures afférentes au Système I, elles sont supposées évoluer continuellement au cours de l'apprentissage ou du développement, en fonction de l'ajustement progressif du Système I aux régularités du matériel traité. Or rien de tel ne correspond à cette organisation progressive dans les réseaux actuels. Les entrées restent identiques à elles-mêmes (par exemple, des syllabes) tout au long de l'apprentissage. On peut évidemment imaginer

des réseaux en cascade, la sortie de l'un servant d'entrée à l'autre, mais le résultat est dépourvu de la puissance d'un système bouclant sur lui-même. En revanche, PARSEr est fondé sur des principes dynamiques d'auto-organisation qui se révèlent totalement compatibles avec la structure du système neuronal décrite plus haut. En effet, dans PARSEr, la perception façonne les représentations internes (des primitives perceptives perçues simultanément deviennent de nouvelles représentations), mais en retour, la perception est continuellement guidée par les unités déjà formées (les nouvelles représentations deviennent de nouvelles primitives perceptives lorsqu'elles sont suffisamment cohésives).

Conclusion: le rôle de la modélisation

Dans cette présentation sélective du champ de la mémoire et de l'apprentissage, nous avons donné beaucoup d'importance à l'approche computationnelle. Ce choix est trompeur à certains égards: en fait, considérée en proportion de l'ensemble des travaux du domaine, l'approche computationnelle reste assez marginale. Dans les ouvrages de synthèse sur la mémoire et l'apprentissage (Byrne 2008; Tulving et Craik 2000), les auteurs des quelques chapitres consacrés à la modélisation commencent généralement par noter la pauvreté relative de leur sujet et par s'interroger sur le pourquoi de cette désaffection. Une hypothèse est que l'intérêt réel de la modélisation reste mal compris. Le propre de la modélisation est souvent pensé comme étant de nature à contraindre à une formulation plus précise des hypothèses, l'objectif étant d'accroître la finesse des prédictions qui peuvent en être déduites. La possibilité de comparaisons *quantitatives* entre les prédictions du modèle et les comportements observés apparaît comme la finalité ultime. C'est peut-être le cas dans certains domaines. Mais les données présentées dans ce chapitre suggèrent qu'il en va différemment dans le champ de la mémoire et de l'apprentissage. La recherche de précision quantitative ne s'est jamais révélée très fructueuse. Par exemple, les nombreux travaux visant à déterminer si la courbe d'oubli s'ajuste mieux à telle ou telle fonction mathématique (par exemple, logarithmique, puissance ou hyperbolique) n'ont pas conduit à des conclusions claires. La seule

conclusion faisant l'objet d'un consensus est celle d'une fonction négativement accélérée, telle que l'avait déjà décrite Ebbinghaus en 1885 : l'oubli, rapide au début, tend à s'atténuer au fil du temps (pour une synthèse, cf. Nairne et Pandeirada 2008).

L'intérêt majeur de la modélisation, nous semble-t-il, est bien plutôt d'examiner la viabilité d'hypothèses alternatives au mode de pensée dominant. Comme l'histoire de la psychologie cognitive l'illustre abondamment, la première explication qui vient à l'esprit face à la découverte d'un nouveau phénomène par les psychologues expérimentalistes ou les neuropsychologues manque bien souvent de parcimonie. La plupart du temps, il s'agit de postuler l'existence de processus spécialisés, *ad hoc*, en correspondance directe avec les comportements observés. Ainsi, comme nous l'avons vu, l'observation de dissociations neuropsychologiques ou comportementales évoque-t-elle naturellement l'existence de systèmes de mémoires indépendants. À un autre niveau, le respect de la syntaxe du langage par l'enfant évoque l'existence d'un analyseur syntaxique autonome, et si les capacités requises pour l'acquisition semblent dépasser les capacités d'analyse de l'enfant, la solution de l'innéisme vient immédiatement se substituer à l'apprentissage en guise d'explication. Remettre en cause un mode de pensée apparemment cohérent avec un grand nombre d'observations est une démarche longue et difficile. La difficulté vient principalement du fait que les explications alternatives sont souvent plus difficiles à comprendre que l'explication de départ. Ainsi, face à une dissociation comportementale, imaginer que chaque comportement est le produit d'un système particulier ne réclame aucun effort de compréhension, ce qui n'est pas le cas d'une explication fondée sur un système unique. De la même façon, imaginer que la maîtrise de la syntaxe repose sur l'existence d'un analyseur syntaxique est plus simple à comprendre qu'une explication basée sur l'ajustement du système aux propriétés statistiques d'événements corrélés perceptivement saillants.

Même si, *du point de vue de l'apprenant*, le gain en parcimonie est considérable, le coût cognitif pour le chercheur est indiscutablement plus élevé. Face à cette situation, montrer qu'un modèle computationnel reposant sur des postulats élémentaires, parfaitement bien identifiés, est suffisant pour rendre compte des données apparaît comme une stratégie de recherche irremplaçable pour montrer que d'autres

systèmes d'explication sont possibles. Évidemment, les conclusions d'une telle démarche peuvent également être celles d'une *impossibilité*, comme nous l'avons vu à propos de l'existence de systèmes de mémoires complémentaires (McClelland *et alii* 1995). Mais la finalité est la même, celle d'explorer des possibilités. Pour remplir cette fonction indispensable dans la dynamique de la recherche sur la mémoire et l'apprentissage, on ne peut que souhaiter l'avènement d'une meilleure intégration de l'approche computationnelle avec les approches plus empiriques de la psychologie cognitive et de la neuropsychologie.

Références bibliographiques

- Anderson, J. R., 1983. *The Architecture of Cognition*, Cambridge (Mass.), Harvard University Press (Cognitive Science Series; 5).
- Berry, C. J., Shanks, D. R. et Henson, R. N. A., 2008. « A unitary signal-detection model of implicit and explicit memory », *Trends in Cognitive Sciences*, 12 (10) : 367-373.
- Blaxton, T. A., 1989. « Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework », *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15 (4) : 657-668.
- Brent, M. R., 1996. « Advances in the computational study of language acquisition », *Cognition*, 61 (1-2) : 1-38.
- Brent, M. R. et Cartwright, T. A., 1996. « Distributional regularity and phonotactic constraints are useful for segmentation », *Cognition*, 61 (1-2) : 93-125.
- Brown, G. D. A., Della Sala, S., Foster, J. K. et Vousden, J. I., 2007. « Amnesia, rehearsal, and temporal distinctiveness models of recall », *Psychonomic Bulletin and Review*, 14 (2) : 256-260.
- Brown, G. D. A., Neath, I. et Chater, N., 2007. « A temporal ratio model of memory », *Psychological Review*, 114 (3) : 539-576.
- Byrne, J. H. (éd.), 2008. *Learning and Memory: A Comprehensive Reference*, 4 volumes, Oxford, Elsevier.

- Christiansen, M. H., Allen, J. et Seidenberg, M. S., 1998. « Learning to segment speech using multiple cues: A connectionist model », *Language and Cognitive Processes*, 13 (2-3): 221-268.
- Cohen, N. J. et Squire, L. R., 1980. « Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that », *Science*, 210 (4466): 207-209.
- Crowder, R. G., 1993. « Short-term memory: Where do we stand? » *Memory and Cognition*, 21 (2): 142-145.
- Ebbinghaus, H., 1885. *Über das Gedächtnis*, Leipzig, Dunker und Humblot.
- Elman, J. L., 1990. « Finding structure in time », *Cognitive Science*, 14 (2): 179-211.
- Fiser, J. et Aslin, R. N., 2001. « Unsupervised statistical learning of higher-order spatial structures from visual scenes », *Psychological Science*, 12 (6): 499-504.
- Fiser, J. et Aslin, R. N., 2005. « Encoding multielement scenes: Statistical learning of visual feature hierarchies », *Journal of Experimental Psychology: General*, 134 (4): 521-537.
- Frank, M. C., Goldwater, S., Griffiths, T. L. et Tenenbaum, J. B., 2010. « Modeling human performance in statistical word segmentation », *Cognition*, 117 (2): 107-125.
- Giroux, I. et Rey, A., 2009. « Lexical and sublexical units in speech perception », *Cognitive Science*, 33 (2): 260-272.
- Goldwater, S., Griffiths, T. L. et Johnson, M., 2009. « A Bayesian framework for word segmentation: Exploring the effects of context », *Cognition*, 112 (1): 21-54.
- Hintzman, D. L., 1986. « "Schema abstraction" in a multiple-trace memory model », *Psychological Review*, 93 (4): 411-428.
- Jusczyk, P., 1997. *The Discovery of Spoken Language*, Cambridge (Mass.), The MIT Press (Bradford Books).
- Keane, M. M., Gabrieli, J. D. E., Noland, J. S. et McNealy, S. I., 1995. « Normal perceptual priming of orthographically illegal nonwords in amnesia », *Journal of the International Neuropsychological Society*, 1 (5): 425-433.
- Kinder, A. et Shanks, D. R., 2003. « Neuropsychological dissociations between priming and recognition: A single-system connectionist account », *Psychological Review*, 110 (4): 728-744.
- McClelland, J. L., McNaughton, B. L. et O'Reilly, R. C., 1995. « Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory », *Psychological Review*, 102 (3): 419-457.
- McClelland, J. L. et Patterson, K., 2002. « Rules or connections in past-tense inflections: What does the evidence rule out? » *Trends in Cognitive Sciences*, 6 (11): 465-472.
- Monaghan, P., Chater, N. et Christiansen, M. H., 2005. « The differential role of phonological and distributional cues in grammatical categorisation », *Cognition*, 96 (2): 143-182.
- Nadel, L., 2008. « Multiple memory systems: A new view », in J. H. Byrne (éd.), *Learning and Memory: A Comprehensive Reference*, vol. 1, R. Menzel (éd.), *Learning Theory and Behaviour*, Oxford, Elsevier: 41-52.
- Nairne, J. S. et Pandeirada, J. N. S., 2008. « Forgetting », in J. H. Byrne (éd.), *Learning and Memory: A Comprehensive Reference*, vol. 2, H. L. Roediger III (éd.), *Cognitive Psychology of Memory*, Oxford, Elsevier: 179-194.
- O'Reilly, R. C. et Rudy, J. W., 2001. « Conjunctive representations in learning and memory: Principles of cortical and hippocampal function », *Psychological Review*, 108 (2): 311-345.
- Pacton, S., Fayol, M. et Perruchet, P., 2005. « Children's implicit learning of graphotactic and morphological regularities in French », *Child Development*, 76 (2): 324-339.
- Perruchet, P., 2008. « Implicit learning », in J. H. Byrne (éd.), *Learning and Memory: A Comprehensive Reference*, vol. 2, H. L. Roediger III (éd.), *Cognitive Psychology of Memory*, Oxford, Elsevier: 597-621.
- Perruchet, P. et Tillmann, B., 2010. « Exploiting multiple sources of information in learning an artificial language: Human data and modeling », *Cognitive Science*, 34 (2): 255-285.
- Perruchet, P. et Vinter, A., 1998. « PARSER: A model for word segmentation », *Journal of Memory and Language*, 39 (2): 246-263.
- Postman, L. et Phillips, L. W., 1965. « Short-term temporal changes in free recall », *Quarterly Journal of Experimental Psychology*, 17 (2): 132-138.

- Ramscar, M., 2002. « The role of meaning in inflection: Why the past tense does not require a rule », *Cognitive Psychology*, 45 (1): 45-94.
- Ranganath, C. et Blumenfeld, R. S., 2005. « Doubts about double dissociations between short- and long-term memory », *Trends in Cognitive Sciences*, 9 (8): 374-380.
- Redington, M., Chater, N. et Finch, S., 1998. « Distributional information: A powerful cue for acquiring syntactic categories », *Cognitive Science*, 22 (4): 425-469.
- Robinet, V. et Lemaire, B., 2009. « MDLChunker: A MDL-based model of word segmentation », in N. Taatgen et H. van Rijn (éd.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society, Amsterdam, July 29-August 1, 2009*, Austin (Tex.), Cognitive Science Society: 2866-2871.
- Rumelhart, D. E. et McClelland, J. L., 1986. « On learning the past tenses of English verbs », in J. L. McClelland, D. E. Rumelhart et the PDP Research Group (éd.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 2, *Psychological and Biological Models*, Cambridge (Mass.), The MIT Press (Bradford Books): 216-271.
- Saffran, J. R., Aslin, R. N. et Newport, E. L., 1996. « Statistical learning by 8-month-old infants », *Science*, 274 (5294): 1926-1928.
- Saffran, J. R., Newport, E. L. et Aslin, R. N., 1996. « Word segmentation: The role of distributional cues », *Journal of Memory and Language*, 35 (4): 606-621.
- Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A. et Barrueco, S., 1997. « Incidental language learning », *Psychological Science*, 8 (2): 101-105.
- Scoville, W. B. et Milner, B., 1957. « Loss of recent memory after bilateral hippocampal lesions », *Journal of Neurological and Neurosurgical Psychiatry*, 20 (1): 11-21.
- Shallice, T. et Warrington, E. K., 1970. « Independent functioning of verbal memory stores: A neuropsychological study », *Quarterly Journal of Experimental Psychology*, 22 (2): 261-273.
- Sherry, D. F. et Schacter, D. L., 1987. « The evolution of multiple memory systems », *Psychological Review*, 94 (4): 439-454.
- Tulving, E., 1972. « Episodic and semantic memory », in E. Tulving et W. Donaldson (éd.), *Organization of Memory*, New York, Academic Press: 381-403.
- Tulving, E., 2007. « Are there 256 different kinds of memory? » in J. S. Nairne (éd.), *The Foundations of Remembering: Essays in Honor of Henry L. Roediger III*, New York, Psychology Press: 39-52.
- Tulving, E. et Craik, F. I. M., 2000. *The Oxford Handbook of Memory*, New York, Oxford University Press.
- Turk-Browne, N. B., Jungé, J. A. et Scholl, B. J., 2005. « The automaticity of visual statistical learning », *Journal of Experimental Psychology: General*, 134 (4): 552-564.
- Warrington, E. K. et Weiskrantz, L., 1968. « New method of testing long-term retention with special reference to amnesic patients », *Nature*, 217 (5132): 972-974.
- Woollams, A. M., Joanisse, M. et Patterson, K., 2009. « Past-tense generation from form versus meaning: Behavioural data and simulation evidence », *Journal of Memory and Language*, 61 (1): 55-76.