

une liste du type: chien, chat, cheval, dauphin, baleine, cachalot... Des termes contigus ont presque toujours un lien. Ils peuvent tous deux désigner des animaux de compagnie, des représentants de la faune marine, des animaux rares dans la vie quotidienne (tels des reptiles ou des insectes)... Pour élaborer une liste, le cerveau recourt à des associations d'idées liées sémantiquement: le choix de l'animal suivant est déterminé par sa similarité avec le précédent selon des propriétés diverses (telles que la taille, l'habitat ou la familiarité).

Des catégories et des « animaux ponts »

Cette expérience a été menée en 2010 par une équipe de l'Université de Navarre, en Espagne, dirigée par Joaquín Goñi et Pablo Villoslada, avec la collaboration de l'un d'entre nous (B. Corominas-Murtra). En la répétant maintes fois avec de nombreux participants, nous avons établi une liste de mots de plus en plus longue et tracé un réseau qui prenait en compte leurs liens sémantiques (voir l'encadré page ci-contre). Ce réseau était organisé en groupes d'éléments ayant davantage de connexions entre eux qu'avec le reste du réseau. Ces groupes sont, par exemple, les insectes, les reptiles, les animaux rencontrés dans les zoos...

Cette structure reflète une organisation fondée sur des catégories. Elle révèle aussi l'existence, d'un point de vue sémantique, d'« animaux ponts », qui permettent de passer d'un groupe à un autre. Ainsi, le crocodile est perçu à la fois comme un reptile et comme un animal de zoo. Le maillage résultant est très interconnecté et facile à parcourir.

De tels réseaux renseignent aussi sur l'origine de certains types d'ambiguïtés sémantiques, apparemment universels. Ainsi, dans toutes les langues, il existe de nombreux termes ayant la même signification et, à l'inverse, des mots qui ont plusieurs sens. Les premiers sont les synonymes et les seconds sont qualifiés de polysémiques.

Le phénomène de la polysémie est surprenant. Le langage tend à favoriser une communication efficace et donc, en principe, à éviter l'ambiguïté. Pourquoi alors la polysémie est-elle omniprésente ? Si les mots polysémiques ne faisaient qu'entraver la communication, la constante évolution des langues (bien plus rapide que l'évolution génétique) les aurait éliminés.

En 2002, les physiciens argentins Mariano Sigman et Guillermo Cecchi ont analysé l'un des réseaux d'association sémantique les plus complets au monde: *WordNet*. Son édification a démarré en 1985 à l'Université de Princeton, aux États-Unis, sous l'impulsion du linguiste George Miller. M. Sigman et G. Cecchi ont utilisé les outils de la théorie des réseaux complexes pour analyser l'organisation globale de *WordNet*.

Confrontés aux mots *lion* et *rayures*, nous pensons souvent au mot *tigre*, qui apparaît comme une façon d'associer ces deux termes *a priori* sans rapport. Ainsi, en choisissant deux mots au hasard, nous pouvons les relier par une chaîne de vocables intermédiaires. Des chemins de mots, de longueur mesurable, apparaissent alors. De même, les liens triangulaires, où deux éléments liés à un troisième sont aussi associés l'un à l'autre, fournissent une idée du degré d'association locale entre concepts: par exemple, les mots *lumière* et *nuit* sont reliés à *lampe*, et aussi entre eux. Les connexions entre les divers nœuds sont de nature distincte, mais on peut les considérer simultanément pour se représenter le réseau comme un tout organisé en différents niveaux.

Le langage humain, un petit monde

M. Sigman et G. Cecchi ont montré que *WordNet* est un réseau de type petit monde, c'est-à-dire où, d'une part, deux nœuds quelconques peuvent être reliés par un chemin assez court et, d'autre part, certaines zones comprennent un grand nombre de connexions par rapport au nombre de nœuds. Le phénomène du petit monde a été analysé en détail en 1998 par Duncan Watts et Steven Strogatz, de l'Université Cornell, aux États-Unis. Sa découverte remonte à quelques dizaines d'années plus tôt, lorsque des sociologues, tel l'Américain Stanley Milgram, ont étudié la proximité sociale entre individus. Imaginons que chacune des personnes qui constituent une société corresponde à un nœud d'un réseau. Si deux d'entre elles se connaissent, elles sont considérées comme connectées. Le degré de séparation sociale entre deux individus quelconques est défini comme le nombre minimum de « sauts » à effectuer entre les nœuds pour les relier.

À partir d'une série d'expériences fondées sur l'échange de cartes postales,

Milgram a déduit que la distance sociale moyenne entre les habitants des États-Unis s'élève à environ six degrés. Ainsi, même dans un pays comptant plusieurs dizaines de millions d'habitants, nous sommes bien plus proches les uns des autres que nous l'imaginons. Le phénomène du petit monde se manifeste dans de nombreux autres réseaux: gènes, circuits électroniques, etc. Il favorise le transport d'information ou d'énergie dans l'ensemble du réseau.

Bien que le lexique *WordNet* analysé par M. Sigman et G. Cecchi soit composé de 66 000 mots, la distance moyenne entre deux d'entre eux ne dépasse pas sept degrés. On peut donc naviguer aisément dans le réseau. En outre, ce dernier présente de nombreux liens triangulaires (une autre caractéristique des réseaux de type petit monde), ce qui facilite la libre association de concepts.

WordNet est également un réseau dit sans échelle. Ce type de réseau se caractérise par l'association d'un grand nombre d'éléments ayant peu de liens entre eux et de quelques nœuds fortement connectés. En 1999, Albert-László Barabási et Réka Albert, alors à l'Université Notre-Dame, aux États-Unis, ont montré que c'est la présence de tels superconnecteurs qui donne lieu au phénomène de petit monde. Bien que la plupart des nœuds de ces réseaux ne permettent pas de se déplacer avec rapidité, les chemins possibles se multiplient dès qu'un superconnecteur est atteint. Et il est facile, grâce à certains d'entre eux, de parvenir rapidement à destination. En outre, la perte de quelques superconnecteurs provoque une baisse de connectivité importante du réseau.

Dans le réseau sémantique, les superconnecteurs sont les mots polysémiques. M. Sigman et G. Cecchi ont analysé les conséquences de leur élimination: le degré de séparation moyen entre les mots est monté jusqu'à 11 et le nombre de liens triangulaires a été divisé par 300.

Une navigation facilitée

La polysémie confère donc une grande cohésion au réseau sémantique, car elle facilite aussi bien la navigation que l'association locale entre concepts (corrélée au nombre de liens triangulaires). L'ambiguïté favorise les liens sémantiques: c'est peut-être pourquoi toutes les langues du monde présentent de la polysémie. L'exécution de

QUAND LES RÉSEAUX SÉMANTIQUES EXPLIQUENT LA POLYSÉMIE

L'étude quantitative des propriétés émergentes du langage se fonde sur la théorie des réseaux complexes. On relie des mots au sein d'un réseau, dont on analyse les propriétés statistiques.

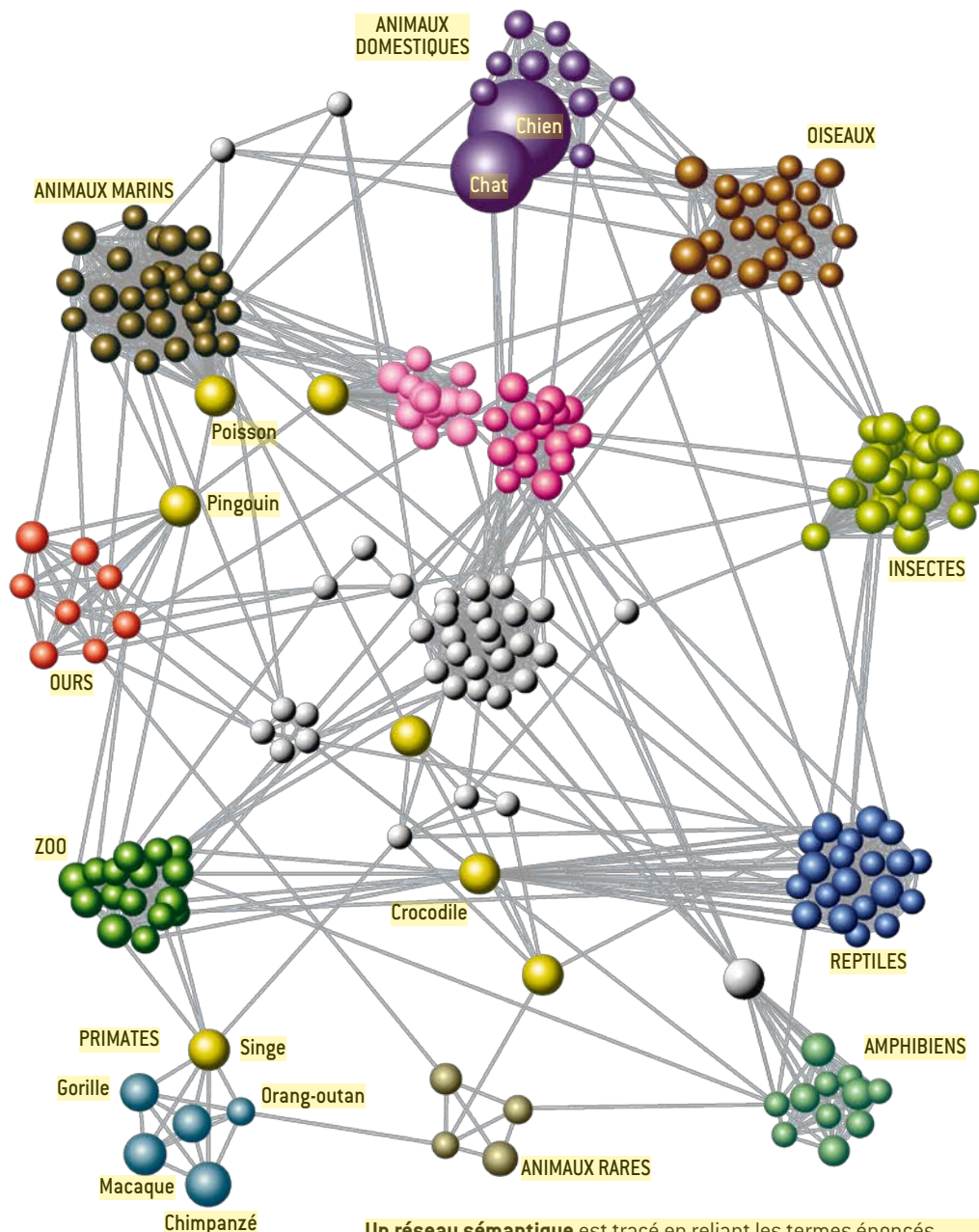
Le schéma ci-dessous illustre une expérience menée en 2010. Les sujets devaient énoncer en un temps court les noms de tous les animaux

qui leur venaient à l'esprit. Deux mots consécutifs avaient presque toujours un lien sémantique. On a alors tracé un réseau de relations sémantiques entre ces termes. On y distingue des groupes de mots (*figurés en couleur*) ayant plus de relations entre eux qu'avec le reste du réseau, et des mots ponts (*nœuds jaunes*), qui mettent en relation di-

vers groupes. Par exemple, le mot *singe* fait penser à la fois aux animaux de zoo et aux primates. Ces caractéristiques facilitent notablement la navigation au sein du réseau, ce qui constitue un avantage pour de multiples tâches cognitives.

Plusieurs études ont montré que les réseaux linguistiques sont sans échelle. Ils se caractérisent par un

petit nombre de nœuds très connectés, ce qui permet d'aller rapidement d'un point à un autre. Dans les réseaux sémantiques, ces « superconnecteurs » seraient les mots polysémiques, ce qui expliquerait leur existence dans toutes les langues (en apparence contre-productive, puisqu'ils introduisent de l'ambiguïté).



Un réseau sémantique est tracé en reliant les termes énoncés successivement, de façon spontanée, par les sujets. La dimension relative des nœuds symbolise la fréquence à laquelle apparaissent les différents mots dans les listes fournies.

Avec l'aimable autorisation des auteurs, source : « The Semantic Organization of the Animal Category : Evidence from Semantic Verbal Fluency and Network Theory », J. Goffi et al. Cognitive Processing, vol. 12, n. 2, pp. 185-196, mai 2011.

tâches cognitives demande de naviguer à travers un vaste univers de mots et de significations, d'où l'utilité du phénomène de petit monde. Des études ultérieures sur d'autres réseaux linguistiques, tels ceux obtenus à partir des dictionnaires de synonymes et d'antonymes, ont confirmé ces propriétés.

Outre ses propriétés sémantiques, l'un des attributs essentiels du langage humain réside dans son pouvoir combinatoire. À partir d'éléments simples et de quelques règles de base permettant de les associer, nous créons d'autres éléments, plus complexes : nous unissons des phonèmes pour construire des syllabes, que nous combinons pour former des mots, avec lesquels nous constituons des phrases. Le nombre de possibilités croît alors de façon exponentielle.

Pour étudier ce potentiel combinatoire, le processus d'acquisition de la langue maternelle offre un champ d'investigation unique. Un bébé n'apprend pas les règles d'une langue comme le ferait un adulte, mais il les construit à partir de données dispersées, fournies par d'autres membres de sa communauté.

L'enfant subit une immersion linguistique nécessairement incomplète, car les mots et les phrases auxquels il est confronté ne peuvent couvrir toutes les éventualités du langage. Il ne reçoit ni une liste d'exemples exhaustive ni des instructions explicites quant à l'usage des règles linguistiques. Le locuteur se forge une connaissance tacite de la langue. Il emploie diverses règles, notamment grammaticales, mais n'est pas conscient de la structure sous-jacente à ses propres productions linguistiques.

Des premiers mots aux premières phrases

Au cours des deux premières années de sa vie, l'enfant entend souvent certains mots et quelques séquences courantes composées de deux vocables, ainsi qu'un faible nombre de phrases (comparé à l'univers des combinaisons possibles). Il passe alors par une phase durant laquelle il n'utilise que des mots isolés, puis par l'étape dite des deux mots, où il associe deux termes de façon assez primitive. Vers l'âge de deux ans, une transition se produit : l'enfant commence à construire des phrases complexes, en faisant peu d'erreurs, voire aucune. Ainsi, l'étape des trois mots n'existe pas. Une fois dépassée celle des deux mots, la

LES AUTEURS



Ricard SOLÉ est chercheur à l'Université Pompeu Fabra de Barcelone, en Espagne, et membre de l'Institut de Santa Fe, aux États-Unis.



Bernat COROMINAS MURTRA est chercheur au Département de science des systèmes complexes de l'Université médicale de Vienne, en Autriche.



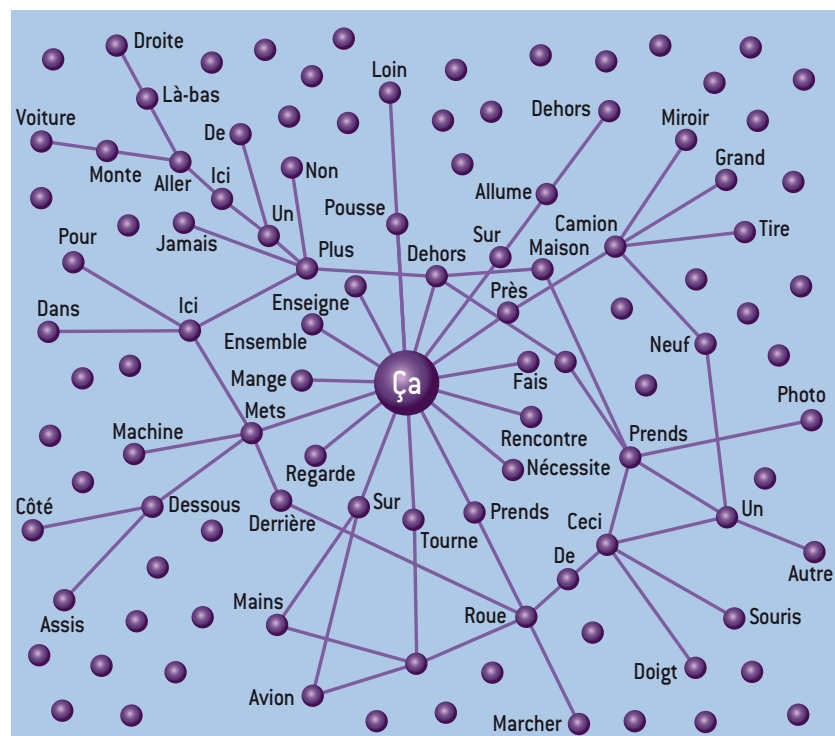
Jordi FORTUNY est chercheur à l'Université de Barcelone.

syntaxe infantile contient déjà une grande partie de la complexité de la syntaxe adulte. Comment l'enfant peut-il développer une connaissance aussi systématique et élaborée à partir d'une stimulation aussi partielle (comprenant très peu d'exemples) et souvent truffée d'ambiguïtés ?

Le processus d'acquisition de la langue a donné lieu à de vifs débats. Dépend-il d'un système cognitif prédisposé à cette acquisition ? En d'autres termes, existe-t-il un « organe du langage » qui facilite le passage d'une communication fragmentée à un discours général, comme le soutient Noam Chomsky ? Pour ce linguiste américain et d'autres, l'enfant dispose d'une grammaire universelle innée. Celle-ci le guiderait au cours de l'acquisition de la langue, en restreignant l'ensemble des combinaisons possibles et en lui procurant les éléments fondamentaux d'analyse. Pour apporter quelques éléments au débat, nous avons exploré la façon dont

L'ÉMERGENCE DE LA SYNTAXE

Vers l'âge de deux ans, les enfants passent directement de la phase dite des deux mots (au cours de laquelle ils n'associent que deux vocables de façon assez primitive) à une étape où ils construisent des phrases complètes. Il n'existe pas d'étape à trois mots. Comment cette transition se produit-elle ?



Au départ, le réseau d'associations syntaxiques a une structure essentiellement arborescente. La plupart des mots sont déconnectés ou ne sont reliés que par paires.

Ricard Solé, Bernat Corominas-Murtra et Jordi Fortuny