

« La façade de la pension donne sur un jardinet, en sorte que la maison tombe à angle droit sur la rue Neuve-Sainte-Geneviève, où vous la voyez coupée dans sa profondeur. »

« La maison où s'exploite la pension bourgeoise appartient à madame Vauquer. Elle est située dans le bas de la rue Neuve-Sainte-Geneviève, à l'endroit où le terrain s'abaisse vers la rue de l'Arbalète par une pente si brusque et si rude que les chevaux la montent ou la descendent rarement. Cette circonstance est favorable au silence qui règne dans ces rues serrées entre le dôme du Val-de-Grâce et le dôme du Panthéon, deux monuments qui changent les conditions de l'atmosphère en y jetant des tons jaunes, en y assombrissant tout par les teintes sévères que projettent leurs coupoles. »

« Le père Goriot, vieillard de soixante-neuf ans environ, s'était retiré chez madame Vauquer, en 1813, après avoir quitté les affaires. Il y avait d'abord pris l'appartement occupé par madame Couture, et donnait alors douze cents francs de pension, en homme pour qui cinq louis de plus ou de moins étaient une bagatelle. »

Extrait du Père Goriot,
Honoré de Balzac, 1835



L'ESSENTIEL

- Les diverses langues du monde ont en commun certains modèles d'organisation et des lois statistiques.
- L'étude de la connectivité des réseaux linguistiques permet d'expliquer de nombreux aspects, tels que la polysémie (l'existence de mots ayant plusieurs sens) ou l'organisation en catégories sémantiques.
- La façon dont la syntaxe émerge chez l'enfant et le développement de formes de communication rudimentaires par des systèmes artificiels renseignent sur l'origine du langage.

D'après Ricard Solé et Josepilh Santiago

1. LE LANGAGE HUMAIN EST STRUCTURÉ
à de multiples niveaux. À partir de quelques règles de composition simples, les phonèmes se combinent en syllabes qui, à leur tour, s'associent en mots. Grâce à la syntaxe, les mots forment des phrases. Le nombre d'éléments possibles croît de façon exponentielle.

de vue systémique. Bien que les langues du monde arborent des différences lexicales et grammaticales notables, elles semblent avoir des structures et des lois statistiques communes. Leur analyse renseigne sur l'origine de la complexité du langage. En particulier, nous avons étudié les mécanismes qui régissent l'acquisition de la syntaxe par l'enfant et l'évolution du langage humain vers un système de communication efficace, souple et au pouvoir informatif illimité.

Pour analyser l'architecture du langage, une approche nouvelle et prometteuse s'appuie sur la théorie des réseaux, qui a connu un développement spectaculaire au cours des dix dernières années. L'idée principale consiste à traiter les mots comme des éléments de base (des nœuds du réseau), à déterminer certaines propriétés qui permettent de les relier et à étudier le maillage résultant.

À titre d'exemple, prenons l'ensemble des noms d'animaux et, en un temps limité à quelques secondes, prononçons tous ceux qui nous viennent à l'esprit. Nous obtiendrons

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