

La structure en réseaux du langage

Ricard Solé, Bernat Corominas-Murtra et Jordi Fortuny

Des techniques d'analyse issues des mathématiques et de la physique révèlent les structures du langage et renseignent ainsi sur son origine et son évolution.

Seule l'espèce humaine emploie un système de communication symbolique capable de transmettre une information complexe. Certains animaux semblent avoir maîtrisé quelques éléments rudimentaires d'un tel système, mais leur mode de communication est beaucoup moins développé que notre langage. De même, aucune machine ne parvient à l'imiter pour l'instant. Comment avons-nous développé ce système de communication ?

Comme l'a signalé le linguiste américain Derek Bickerton, le langage ne laisse pas de fossiles. On doit donc recourir à des données indirectes. On étudie, par exemple, les changements culturels liés à l'art ou à la fabrication d'outils, qui traduisent des améliorations cognitives et, suppose-t-on, des facultés de communication. Ces dernières années, la génétique a aussi apporté quelques éléments. On a ainsi découvert un gène noté *FoxP2*, dont la mutation provoque des troubles de la parole. Sa séquence est identique chez l'homme moderne et chez les Néandertaliens, qui maîtrisaient peut-être certains aspects d'une communication orale.

Ces indices corroborent une autre affirmation de D. Bickerton : « Rien ne survient dans le monde sans laisser une trace, aussi ténue ou indirecte soit-elle. » En linguistique, l'analyse de ces traces a donné lieu à de grands débats, souvent très théoriques, concernant leur nature et leur origine (génétique, culturelle ou les deux).

Certaines recherches en linguistique ont récemment fait appel à des outils associés d'habitude aux mathématiques et à la physique statistique, tels que l'analyse de systèmes complexes. Celle-ci porte sur des systèmes composés d'un grand nombre d'éléments et qui présentent des propriétés dites émergentes : les propriétés du système global sont plus riches que la simple sommation des propriétés de chaque composante.

Le langage humain est un système de ce type. De même qu'une fourmière est plus qu'un simple agrégat d'insectes, le langage ne se limite pas à la juxtaposition de mots.

Dans une série de travaux récents, nous avons abordé l'étude quantitative du langage du point

« Naturellement destiné à l'exploitation de la pension bourgeoise, le rez-de-chaussée se compose d'une première pièce éclairée par les deux croisées de la rue, et où l'on entre par une porte-fenêtre. »

la rue

le père

on entre

il avait

monument

droite

toi

sur

donne

rude

deux

homme

jaune

pouvoir

maison

douze

pa

no

fa

si

es

bi

do

de

ti

ma

bu

lo

m

o

a

l

e

f

r

d

i

c

« 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

la maison
un jardinet
elle est
deux monuments
teinte sévère
la ferme
madame Couture
les cheveux
cinq louis
de plus
de moins
les affaires
le bas
une
gauche
entre
rue
pièce
il

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 Josep Lluís 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