

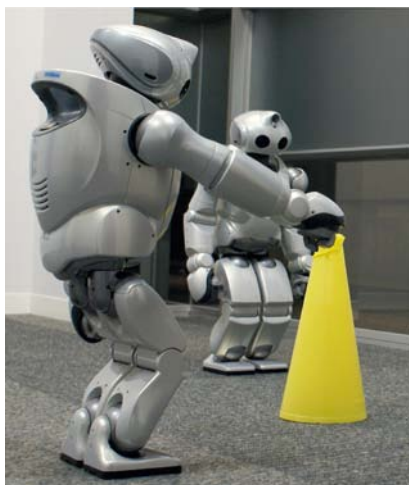
Il n'est efficace pour la communication que grâce à notre capacité d'interpréter les structures produites. Cette interprétation est compositionnelle, c'est-à-dire qu'elle prend en compte les éléments lexicaux d'un énoncé, mais aussi leurs relations dans le contexte de la phrase: ce n'est pas une simple « addition » de significations. La capacité d'interpréter de façon compositionnelle les structures récursives est la clé du potentiel expressif du langage. Pour de nombreux experts, son acquisition au cours de l'évolution est l'une des principales énigmes posées par l'étude des origines du langage.

Le moindre effort

De telles observations restant assez qualitatives, nous avons tenté d'analyser de façon plus quantitative les propriétés des structures syntaxiques. Nous nous sommes intéressés à la fréquence d'usage des mots dans une langue à partir de l'analyse de textes écrits. Elle suit une loi dite de Zipf, en référence au linguiste américain George Zipf qui l'a décrite pour la première fois dans les années 1930. Cette loi stipule que la fréquence d'occurrence $f(r)$ d'un mot est inversement proportionnelle à son rang r (qui désigne sa place dans une liste où les mots sont classés par fréquence d'utilisation décroissante). Autrement dit, si le mot le plus employé (de rang $r = 1$) apparaît un certain nombre de fois dans un texte, le deuxième plus utilisé ($r = 2$) apparaît deux fois moins souvent, le troisième trois fois moins souvent, etc. Il en découle que peu de mots sont très fréquents.

En 2011, nous avons examiné une hypothèse formulée par Zipf au milieu du siècle dernier. Selon cette hypothèse, le langage humain suivrait le principe du moindre effort, qui concernerait aussi bien l'émetteur que le récepteur d'un message. Moins un locuteur emploie de mots, plus il économise son effort, mais plus il augmente celui que son interlocuteur doit faire pour déchiffrer le message. D'après Zipf, le nécessaire compromis qui en résulte pourrait régir l'évolution des langues. Cela expliquerait en particulier qu'aucun lexique n'est composé d'un seul mot (moindre effort pour l'émetteur, qui ne doit apprendre qu'un mot), ni d'innombrables vocables ayant chacun un sens unique (moindre effort pour le récepteur, puisqu'aucun mot n'est ambigu).

Pour vérifier la pertinence de cette hypothèse, nous avons utilisé des résultats



Luc Steels, Laboratoire Sony, Paris

2. LES AUTOMATES de Luc Steels, de la Société Sony CSL, à Paris, apprennent à partir de leur environnement et inventent des mots pour se référer aux objets. En faisant de multiples essais et en communiquant, plusieurs robots peuvent développer un lexique commun et une grammaire simple.

BIBLIOGRAPHIE

B. Corominas-Murtra et al., *Emergence of Zipf's law in the evolution of communication*, *Physical Review E*, vol. 83, art. 036115, 2011.

R. V. Solé et al., *Diversity, competition, extinction : The ecophysics of language change*, *Journal of the Royal Society Interface*, vol. 7, pp. 1647-1664, 2010.

A.-L. Barabási et É. Bonabeau, *Les réseaux invariants d'échelle*, *Dossier Pour la Science*, n°66, 2010.

B. Corominas-Murtra et al., *The ontogeny of scale-free syntax networks : phase transitions in early language acquisition*, *Advances in Complex Systems*, vol. 12, pp. 371-392, 2009.

R. V. Solé et al., *Language networks: their structure, function and evolution*, *Complexity*, vol. 15, pp. 20-26, 2009.

D. Bassano, *Premiers pas dans le langage*, *Cerveau&Psycho*, n°3, 2003.

antérieurs qui quantifiaient les efforts du locuteur et du récepteur grâce à la théorie de l'information. En outre, nous avons supposé que le langage se complexifie en incorporant des éléments nouveaux et qu'il est doté d'une capacité informative illimitée. Nous avons alors montré que le système de communication le moins coûteux en termes d'efforts se conforme à la loi de Zipf. On retrouve donc bien une caractéristique des langues naturelles à partir de l'hypothèse du moindre effort.

Enfin, nous avons établi qu'il est possible de construire un système formel d'emboîtement de contenus où le passage de l'étape à deux éléments à celle à trois éléments implique une transition vers un potentiel de création illimité. Ce système fait intervenir la loi de Zipf, qui associe la précision (de multiples mots peu courants et spécifiques) et l'ambiguïté (peu de termes très fréquents, mais dont l'usage est générique). Si ce cadre théorique est encore abstrait, il suggère que la transition observée dans le processus d'acquisition de la langue obéit en partie à un principe organisationnel fondamental acquis par le cerveau au cours de l'évolution.

Pour étudier les origines de la complexité du langage, on doit aussi prendre en compte le rôle joué par l'évolution culturelle. Les langues subissent des changements notables en un temps bien plus court que celui de l'évolution génétique. Un texte écrit dans notre langue maternelle il y a à peine quelques centaines d'années contient de nombreux archaïsmes et peut être ardu à comprendre. L'entrelacement des évolutions linguistiques et culturelles est difficile à tester expérimentalement, mais la simulation informatique a permis de l'explorer.

Des langages artificiels

Dans le sillage du linguiste britannique Jim Hurford, divers chercheurs, tels Simon Kirby, de l'Université d'Édimbourg, ou Morten Christiansen, de l'Université Cornell, ont mené des expériences virtuelles au cours desquelles une population d'« agents », capables d'échanger et de décoder des mots, ont développé un certain nombre de propriétés caractéristiques du langage. Dans plusieurs d'entre elles, certaines règles d'ordonnement des mots sont apparues de façon spontanée. Selon plusieurs scientifiques, tels S. Kirby et Terrence Deacon, de l'Université de Californie, le langage