

cela, nous avons mesuré le débit syllabique (le nombre de syllabes prononcées par seconde) et estimé la densité syllabique d'information (la quantité d'information portée par chaque syllabe). La notion de quantité d'information utilisée ici est celle définie par le mathématicien américain Claude Shannon au milieu du XX^e siècle.

Au cours de ses travaux sur les télécommunications durant la Seconde Guerre mondiale, Shannon a montré que l'on peut estimer mathématiquement la quantité moyenne d'information transmise entre un émetteur et un récepteur, indépendamment de la signification du message. Dans le cas d'un discours constitué d'une suite de syllabes, on considère en première approche que l'auditeur ignore à chaque instant la syllabe qui sera prononcée ensuite. La quantité moyenne d'information transmise par syllabe est alors définie comme la « quantité d'incertitude » que l'énoncé de cette syllabe vient combler. Plus il y a de syllabes dans une langue, plus la probabilité de prédire la syllabe correcte en répondant au hasard est faible, et donc plus l'incertitude est grande et plus chaque syllabe porte d'information. Une syllabe prononcée en japonais (parmi les 416 possibles) porte donc moins d'infor-

mation au sens de Shannon qu'une syllabe anglaise (issue d'un inventaire de près de 8000 syllabes).

Quantifier l'information

En réalité, d'autres facteurs interviennent, un énoncé n'étant pas constitué de syllabes tirées au hasard. Les syllabes forment des mots : si quelqu'un prononce *an-ti-con-sti-tu-tio*, la syllabe suivante sera à coup sûr *-nel-* et le fait de l'entendre apporte peu d'information. En outre, les mots obéissent à des règles grammaticales. Enfin, les syllabes sont plus ou moins fréquentes dans une langue donnée. En français, par exemple, on utilise bien plus souvent la syllabe qui constitue le mot *et* que celle qui correspond à *strict*. On traduit alors la densité syllabique d'information par une valeur numérique en recourant au formalisme mathématique de la théorie de l'information.

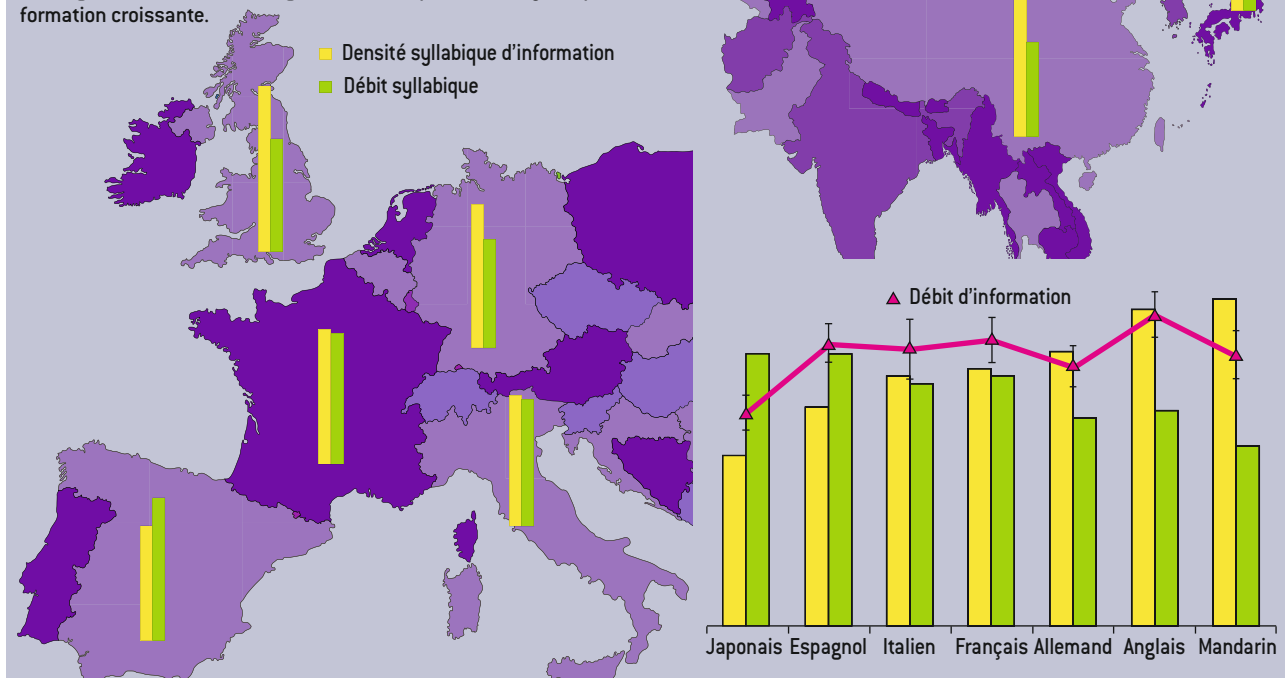
Une syllabe japonaise étant moins informative qu'une syllabe anglaise, on s'attend à ce qu'une même histoire nécessite bien plus de syllabes en japonais qu'en anglais. Là encore, nous l'avons vérifié, avec une vingtaine de courts textes tra-

duits dans chacune des langues de l'étude (voir la figure 3). Ainsi, les versions japonaises comptaient en moyenne 85 pour cent de syllabes supplémentaires par rapport à leurs équivalents anglais.

Faut-il en déduire qu'une même conversation durera près de deux fois plus longtemps en japonais qu'en anglais ? Cela paraît peu probable : toutes les langues du monde étant adaptées à leur fonction d'outil de communication, on imagine difficilement qu'un échange verbal prenne un temps très différent d'une langue à l'autre. Une langue où crier « Au secours ! » prendrait une minute plutôt que quelques fractions de seconde remplirait assez mal sa fonction !

Pour le confirmer, nous avons analysé des enregistrements des 20 textes lus par différents locuteurs. Nous avons montré que les variations de débit syllabique compensent les différences de quantité d'information portée par chaque syllabe : une langue telle que le japonais, qui nécessite plus de syllabes que l'anglais pour véhiculer la même information, a un débit syllabique plus rapide ; la densité d'information moindre est contrebalancée par une vitesse de transmission supérieure. Ainsi, entre les sept langues étudiées, la densité syllabique d'information et le débit syllabique

2. LE DÉBIT D'INFORMATION (triangles rouges sur le graphique en bas à droite) est assez constant pour les langues étudiées. Il est fonction de la densité syllabique d'information (barres jaunes) et du débit syllabique (barres vertes), qui tendent à évoluer en sens contraire. Sur le graphique, les langues sont ordonnées de gauche à droite par densité syllabique d'information croissante.



François Pellegrino, Christophe Coupé et Egidio Marsico

varient en sens contraire, avec pour conséquence un débit d'information très voisin pour la plupart des langues – même si l'anglais et le japonais se démarquent (leurs débits d'information sont respectivement supérieur et inférieur à la moyenne).

Précisons qu'on parle ici des valeurs moyennes. Des différences de débit d'information demeurent à toutes les échelles : chez un locuteur confronté à diverses situations, entre individus, entre populations... Un individu peut ainsi s'adapter au contexte, non seulement en parlant plus ou moins vite, mais également en choisissant les mots employés, les tournures de phrases, etc. Les locuteurs d'une même langue présentent également des différences de débit « normal » d'élocution, certaines personnes parlant naturellement plus vite que d'autres. Dans notre étude, par exemple, les locuteurs français les plus rapides avaient un débit comparable à celui des locuteurs japonais les plus lents.

Ni trop vite ni trop lentement...

Pourquoi le débit moyen d'information ne varie-t-il pas dans des proportions plus importantes ? Un débit trop faible serait socialement pénalisant, car des interactions efficaces nécessitent l'échange d'informations. À l'inverse, un débit moyen trop rapide aurait des coûts physiologique et cognitif trop élevés. En effet, si chacun peut accélérer momentanément son débit de parole (et donc son débit d'information), cela requiert à la fois un supplément d'énergie biomécanique de la part du locuteur et un supplément de ressources neurocognitives de la part de l'auditeur. Plusieurs travaux récents suggèrent d'ailleurs que le cerveau traite plus efficacement les informations présentées à un rythme compris dans une plage étroite, voisine de celle que nous avons observée ; ils se fondent pour cela sur l'analyse des réseaux neuronaux mobilisés par l'audition et la compréhension du langage.

Nous poursuivons actuellement nos recherches en augmentant le nombre de langues étudiées (nous avons ajouté le turc et le coréen, ainsi que plusieurs langues d'Afrique subsaharienne et de Chine), avec l'intention de relier la densité d'information

Hier soir, j'ai ouvert la porte d'entrée pour laisser sortir le chat. La nuit était si belle que je suis descendu dans la rue prendre le frais.
J'avais à peine fait quelques pas que j'ai entendu la porte claquer derrière moi. J'ai réalisé, tout d'un coup, que j'étais enfermé dehors. Le comble c'est que je me suis fait arrêter alors que j'essayais de forcer ma propre porte !

昨夜、私は猫を外に出してやるために玄関を開けてみると、あまりに気持ちのいい夜だったので、新鮮な空気を吸おうと、ついふらっと庭へ降りたのです。すると後ろでドアが閉まって、カチッと言う音が聞こえ、自分自身を締め出してしまったことに気が付いたのです。挙げ句の果てに、私は無理矢理ドアをこじ開けようとしているところを逮捕されてしまったのです。

Last night I opened the front door to let the cat out. It was such a beautiful evening that I wandered down the garden for a breath of fresh air. Then I heard a click as the door closed behind me. I realised I'd locked myself out. To cap it all, I was arrested while I was trying to force the door open !

3. CE TEXTE est l'un de ceux qui ont été traduits dans plusieurs langues pour tester les différences de densité syllabique d'information, de débit syllabique et de débit d'information. En japonais, langue où chaque syllabe porte peu d'information (la densité syllabique est faible), il comporte près de deux fois plus de syllabes qu'en anglais.

à la complexité structurelle des systèmes linguistiques. Les questions sont multiples : quelles différences observe-t-on entre les langues où les mots sont souvent polysyllabiques et celles où ils sont plutôt monosyllabiques ? Comment la densité syllabique d'information varie-t-elle entre les langues aux schémas syllabiques très simples (CV ou CVC par exemple) et celles où ces schémas sont plus complexes (CCVCC par exemple) ? Etc.

Pour comparer les langues, il est également intéressant d'aller au-delà de la notion mathématique d'information que nous avons considérée, pour se pencher sur la notion plus générale de signification. Si les informations sont transmises à des débits comparables d'une langue à l'autre, des différences notables existent quant à ce qui est considéré comme une information pertinente à transmettre explicitement. Le mandarin et le japonais, par exemple, recourent à des éléments grammaticaux nommés classificateurs. Ceux-ci spécifient la forme, la taille ou d'autres caractéristiques physiques ou fonctionnelles des éléments dénombrés. Ainsi, la phrase « Trois chats boivent du lait » dite en japonais se traduit littéralement en français par « Trois animaux-chats du lait boivent ». Sur un autre plan, en japonais, les formules de politesse ont un rôle social très codifié et font partie de l'information jugée essentielle à transmettre lors de la communication.

Nos résultats soulignent qu'il existe plusieurs profils de langues. Ainsi, le mandarin se fonde sur des syllabes denses en information (notamment grâce aux tons) transmises à un débit assez lent, tandis que l'espagnol utilise des syllabes moins denses en information, mais émises plus vite. Cependant, le fait qu'une langue soit plus rapide ou plus dense en information ne la rend pas plus efficace pour communiquer, puisque le débit d'information reste quasi constant. ■

■ BIBLIOGRAPHIE

F. Pellegrino *et al.*,
A cross-language perspective on speech information rate,
Language, vol. 87, pp. 539-558,
2011.

J. Vaissière, **La phonétique**,
Collection Que sais-je ?, Presses
Universitaires de France, 2006.

Bernard Comrie *et al.*, **Atlas des langues : L'origine et le développement des langues dans le monde**, Acropole, 2004.

Physique théorique

Boucles et arbres, à la recherche d'une nouvelle physique

Zvi Bern, Lance Dixon et David Kosower

On calcule les probabilités des processus impliquant les particules élémentaires à l'aide de diagrammes dits de Feynman. Une nouvelle méthode simplifie ces calculs et remet au goût du jour certaines théories unificatrices.

