

Les langues du monde : un même débit d'information

Le débit de parole varie selon les langues, mais plus il est rapide, moins chaque syllabe véhicule d'information. Ainsi, quelle que soit la langue, la vitesse de transmission de l'information est quasi constante.

François Pellegrino, Christophe Coupé et Egidio Marsico

A la terrasse d'un café, dans le brouhaha ambiant, votre attention est soudain attirée par un voisin parlant une langue étrangère. Sa voix vous semble traînante ou, à l'inverse, empressée. Cette impression, qu'on pourrait croire subjective, correspond à une réalité mesurable : le débit syllabique (le nombre de syllabes prononcées par seconde) varie d'une langue à l'autre. Nous l'avons récemment montré en comparant les débits syllabiques d'une soixantaine de locuteurs s'exprimant dans sept langues d'Europe et d'Asie.

L'impression que les Espagnols parlent vite, par exemple, est fondée : en moyenne, ils prononcent 26 pour cent de syllabes en plus par seconde que les locuteurs de l'anglais et 50 pour cent de plus que ceux du chinois mandarin. Pour autant, une même histoire durera-t-elle une fois et demie plus longtemps en mandarin qu'en espagnol ?

Non, car nous verrons que la différence de débit syllabique est compensée par la quantité d'information portée par chacune des syllabes (on parle de densité syllabique d'information) : une langue rapide recourt à un plus grand nombre de syllabes qu'une langue lente pour raconter la même histoire, chaque syllabe portant moins d'information. Finalement, les locuteurs des

différentes langues échangent des informations à un débit comparable. Pour le montrer, nous présenterons d'abord les ensembles de syllabes (les inventaires syllabiques) à disposition des langues et nous précisons la notion d'information dans ce contexte.

Des contraintes façonnant les langues

Au cours du XX^e siècle, les linguistes ont étudié plusieurs centaines de langues, en insistant tantôt sur leurs différences et tantôt sur leurs ressemblances. En 1957, le linguiste américain Martin Joos a affirmé que les langues peuvent différer les unes des autres sans aucune limite et de façon imprévisible. À l'inverse, Noam Chomsky a supposé qu'elles sont toutes sous-tendues par une même grammaire innée et universelle, dont elles seraient des reflets déformés. Plus récemment, la linguistique dite fonctionnelle a adopté un point de vue différent : elle décrit à la fois la diversité des langues et les tendances qu'elles partagent comme résultant de la compétition entre des contraintes variées. Celles-ci peuvent être physiologiques (imposées par le système vocal), cognitives (issues notamment de la capacité du cerveau à traiter l'information), communicatives (on

L'ESSENTIEL

- Grâce à la mesure du débit syllabique (le nombre de syllabes prononcées par seconde) et de la densité syllabique d'information (la quantité d'information portée par chaque syllabe), on évalue le débit d'information de différentes langues.
- Ce débit est presque constant d'une langue à l'autre. Il correspondrait à une vitesse de transmission adaptée aux besoins de la communication.
- Pour l'instant, seules sept langues ont été étudiées. Ce nombre doit être augmenté pour confirmer et étendre les résultats.

1. QUELQUE 6 000 LANGUES sont parlées dans le monde, parfois par des centaines de millions de personnes, tel le mandarin en Chine, ou par quelques centaines de locuteurs, comme le cavineña en Bolivie. Elles semblent aussi efficaces les unes que les autres pour transmettre l'information, bien que le débit de la parole varie notablement.



djem/Shutterstock



Jianhao Guan/Shutterstock

Saim October/Shutterstock



REDAV/Shutterstock



Dima Sobko/Shutterstock

■ LES AUTEURS

François PELLEGRINO, Christophe COUPÉ et Egidio MARSICO travaillent au laboratoire Dynamique du langage (DDL), à l'Institut des sciences de l'homme (ISH), à Lyon.

■ SUR LE WEB

M. Dryer et M. Haspelmath (éds), *The World Atlas of Language Structures Online*, Munich, Max Planck Digital Library 2011: <http://wals.info/>.

doit par exemple parler l'un après l'autre) ou sociales (la parole exprime qui l'on est vis-à-vis des autres, auxquels on s'adresse avec des formes plus ou moins polies, tels les pronoms personnels *tu* et *vous* en français). Elles façonnent en permanence les langues, qui sont, dès lors, des systèmes dynamiques complexes.

Depuis le début du XXI^e siècle, le développement de méthodes d'analyse de ces systèmes, conjugué à la multiplication de données linguistiques disponibles, permet de mener des études quantitatives d'envergure. En particulier, on cherche à déterminer – sans réponse claire pour l'instant – si les langues présentent une complexité égale. Cette question est difficile, car aucune définition de la complexité ne fait consensus en linguistique : certains tentent de la quantifier à partir de la durée d'apprentissage d'une langue par les

enfants, d'autres à partir du nombre d'opérations mentales (déductions, références au contexte, etc.) nécessaires au décodage d'un message... De façon schématique, la complexité d'une langue correspond à la taille du « mode d'emploi » qu'un locuteur doit maîtriser pour s'exprimer correctement dans cette langue.

Nos recherches s'inscrivent dans ces tentatives de « quantifier » les langues, mais ont un but différent : plutôt que de comparer la complexité des langues, nous cherchons à montrer qu'elles offrent la même capacité à communiquer, *via* la mesure du débit d'information.

Cette mesure requiert la compréhension des fondements de notre système de communication : la production, la transmission et la perception de messages linguistiques sonores. Ces messages sont constitués de phrases, décomposables en

LES SONS DES LANGUES DU MONDE

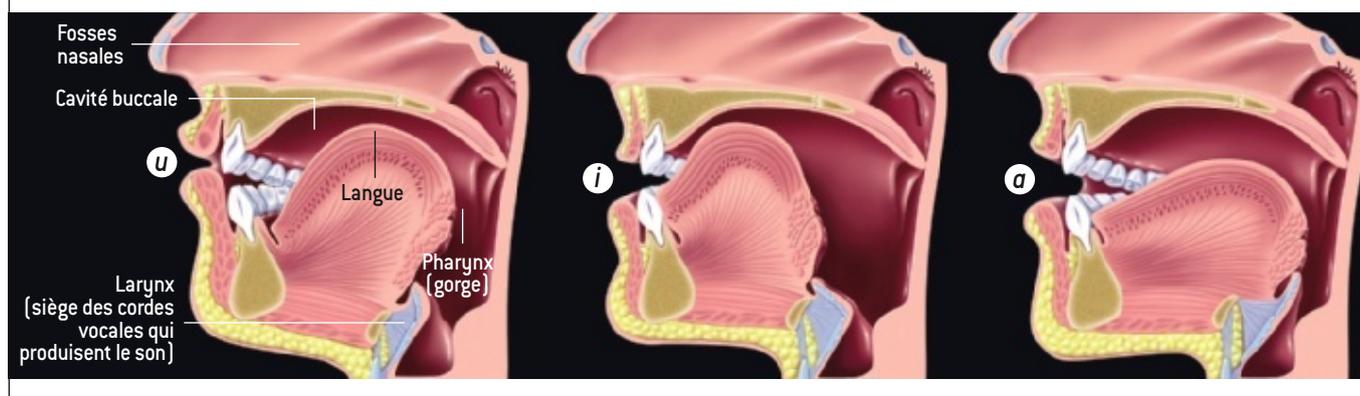
La parole est formée d'enchaînements de sons continus (les voyelles) et de sons « bloqués » (les consonnes). Ce sont les deux éléments de base du langage. Aucune langue n'est constituée uniquement de voyelles ou de consonnes – ce dernier terme signifiant d'ailleurs « qui sonne avec », sous-entendu avec une voyelle. Les sons sont orchestrés par le système vocal. Pour prononcer les voyelles, le système vocal module l'air de différentes façons (langue plus ou moins proche du palais, lèvres arrondies ou non...). Divers gestes articulatoires peuvent augmenter le nombre de nuances. Par exemple, en français, la nasalisation (l'envoi d'air dans les fosses nasales) transforme le son de *a* en *an*, tandis qu'en allemand, la durée de la voyelle permet de distinguer entre les mots *Rate* (avec un *a* long) et *Ratte* (avec un *a* bref), qui signifient respectivement une mensualité et un rat.

Malgré ces différences, la classe des voyelles est assez homogène du point de vue articula-

toire et c'est (presque) toujours le même ensemble de paramètres acoustiques qui permet de les décrire dans les différentes langues. Ces paramètres, nommés formants, correspondent à des accumulations d'énergie dans certaines bandes de fréquences sonores, produites par des gestes articulatoires entraînant des phénomènes de résonance. Ainsi, le [u] est obtenu en concentrant quasiment toute l'énergie sonore dans des fréquences relativement basses (en dessous de 1 000 hertz), alors que pour le [i], l'énergie est répartie entre des basses fréquences (autour de 300 hertz) et des fréquences plus élevées, vers 2 000 et 2 800 hertz.

À l'inverse, les consonnes sont bien plus diversifiées, tant du point de vue articulatoire qu'acoustique. Cependant, toutes sont produites en entravant le passage de l'air, et le degré d'entrave est le premier paramètre de leur classification. On distingue par exemple

les consonnes occlusives (telles *p*, *t*, *k*, *b*, *d*, *g*), créées avec une interruption totale du flux d'air, et les consonnes fricatives (telles *f*, *v*, *s*, *z*), réalisées avec un rétrécissement laissant échapper un peu d'air. Deux autres paramètres différencient les consonnes : le lieu d'articulation, qui précise où a lieu l'entrave dans le système vocal, et le voisement, qui indique l'activité des cordes vocales. Ainsi, le son [k] est produit en interrompant le flux d'air avec le dos de la langue et sans vibration des cordes vocales, tandis que le son [b] est réalisé par une fermeture complète au niveau des lèvres tout en maintenant une vibration des cordes vocales. Comme pour les voyelles, un certain nombre de gestes articulatoires accroissent l'inventaire des consonnes, par exemple en allongeant la durée d'un son. Dans les différentes langues du monde, on a compté près d'un millier de voyelles et de consonnes



mots, eux-mêmes formés de séquences de sons élémentaires nommés phonèmes, qui sont propres à chaque langue. Les phonèmes comprennent des consonnes, des voyelles et parfois des diphthongues (des voyelles complexes dont le timbre se modifie en cours de prononciation, tel [ei] dans le mot anglais *day*). Précisons qu'on parle ici des voyelles et des consonnes phonétiques et non alphabétiques (voir l'encadré ci-dessous) : en français, la voyelle [o] peut ainsi s'écrire o, au, eau...

L'inventaire des phonèmes d'une langue ne suffit pas pour rendre compte de la façon dont elle « sonne », qui dépend aussi de leur agencement en syllabes. Pour ce faire, on doit donc s'intéresser aux syllabes, les éléments d'organisation rythmique de la parole (les « morceaux » en lesquels un locuteur découpe intuitivement les mots dans sa langue natale). Selon

les langues, une même suite de phonèmes peut être intégrée dans une ou plusieurs syllabes : le français autorise ainsi la succession des sons [s] et [t] dans une même syllabe, comme dans le mot *stop*, mais pas le thaï ; en conséquence, les Thaïlandais prononcent souvent le mot *stop* en insérant une voyelle entre [s] et [t].

Les syllabes les plus simples sont constituées d'une unique voyelle. En français, certaines correspondent à des mots fréquents : la préposition *à*, la conjonction de coordination *et*, le nom commun *eau*... Les syllabes peuvent aussi comporter des consonnes, comme par exemple dans *strict*, une syllabe – et un mot – composée de trois consonnes, suivies d'une voyelle, puis de deux consonnes. On note une telle structure syllabique sous la forme CCCVC, où la lettre C symbolise une consonne et la lettre V une voyelle.

Les structures syllabiques varient dans des proportions importantes entre les langues. Ainsi, la plupart des syllabes utilisées en espagnol ont une structure CV ou CVC, alors que l'anglais a une grande diversité syllabique, incluant des syllabes très longues. Par exemple, le mot *strengths* (signifiant *forces*) est transcrit par les phonéticiens [streŋ(k)θs] (θ traduit le son de *th*) et présente une structure CCCVC(C)CC (la parenthèse indique une consonne qui n'est pas toujours prononcée).

Enfin, des syllabes composées de consonnes et de voyelles identiques se distinguent dans certaines langues par des différences de « mélodie ». C'est souvent le cas en Afrique et en Asie, et parfois en Amérique. Là encore, la richesse de ces systèmes est variable : dans certains idiomes mexicains, une même séquence de phonèmes peut comporter jusqu'à huit mélodies distinctes, nommées tons, qui créent huit mots aux significations totalement différentes ! En Europe, la mélodie est parfois aussi utilisée, par exemple en norvégien, mais de façon plus limitée (on parle d'accents tonals et non de tons).

obtenues entre autres grâce à l'ajout d'un trait secondaire (durée, nasalisation...).

Les études ont vite révélé que de nombreux phonèmes (représentés essentiellement par les consonnes et les voyelles phonétiques) se retrouvent dans les différentes langues. Les scientifiques s'accordent aujourd'hui sur le fait que les systèmes sonores des langues sont le fruit de deux contraintes antagonistes. La première concerne la production de la parole : les sons doivent être choisis afin de limiter l'effort articuloire du locuteur. La seconde contrainte porte sur la perception et impose de faciliter la tâche de l'auditeur en créant un contraste suffisant entre les sons à différencier. Il résulte de cette double contrainte que les systèmes comportant peu de sons utilisent préférentiellement des gestes articulatoires simples (telle la modulation de l'ouverture de la bouche ou de la position de la langue), et que lorsqu'il y a plus de phonèmes, des gestes arti-

culatoires plus compliqués sont nécessaires pour assurer une distance perceptuelle suffisante.

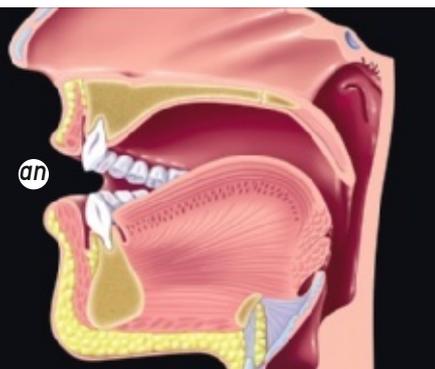
Du fait de ces contraintes structurelles assez fortes – et identiques pour tous –, les langues ont de nombreux points communs. Toutes ont des consonnes occlusives et près de 90 pour cent possèdent le *k*. Plus de 80 pour cent utilisent les voyelles phonétiques *i*, *a*, *ou*, et une majorité sont organisées autour de cinq voyelles de base : *i*, *é*, *a*, *o*, *ou*. Malgré cela, les langues sont très variées. Les systèmes sonores décrits aujourd'hui comportent de 11 à 141 consonnes et voyelles, pour une moyenne de 31 éléments (en moyenne un tiers de voyelles et deux tiers de consonnes). Si la majorité des langues ont moins de huit voyelles, certaines en ont jusqu'à 24, tandis que le nombre de consonnes varie de 6 à 95 ! À titre de comparaison, la langue française comporte moins de 20 consonnes et une quinzaine de voyelles.

20 fois plus de syllabes en anglais qu'en japonais

Chaque langue a son propre nombre de briques élémentaires et des règles d'assemblage spécifiques pour construire ses syllabes. Une langue comme l'anglais, ayant un inventaire de phonèmes étendu et autorisant des syllabes complexes, offre des combinaisons bien plus nombreuses que le japonais, dont les phonèmes sont assez peu nombreux, et qui utilise préférentiellement des syllabes simples.

En analysant ces différents paramètres sur des centaines de milliers, voire de millions, de mots, nous avons comparé les inventaires syllabiques de sept langues (allemand, anglais, espagnol, français, italien, japonais et mandarin). Même sur un échantillon de langues aussi limité, nous observons des variations importantes des tailles des inventaires syllabiques, qui vont de 416 syllabes distinctes en japonais à près de 8 000 en anglais (soit presque 20 fois plus).

Nous nous sommes ensuite intéressés aux conséquences de ces différences sur le débit d'information du discours. Pour



Notre appareil vocal produit des sons variés, grâce à des gestes articulatoires parfois complexes, pouvant impliquer tous ses composants (bouche, gorge, larynx...). De gauche à droite, les configurations représentées correspondent à *u*, *i*, *a* et *an*. Cette dernière voyelle est obtenue en ajoutant au *a* un trait secondaire dit de nasalisation (on ouvre les fosses nasales).

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

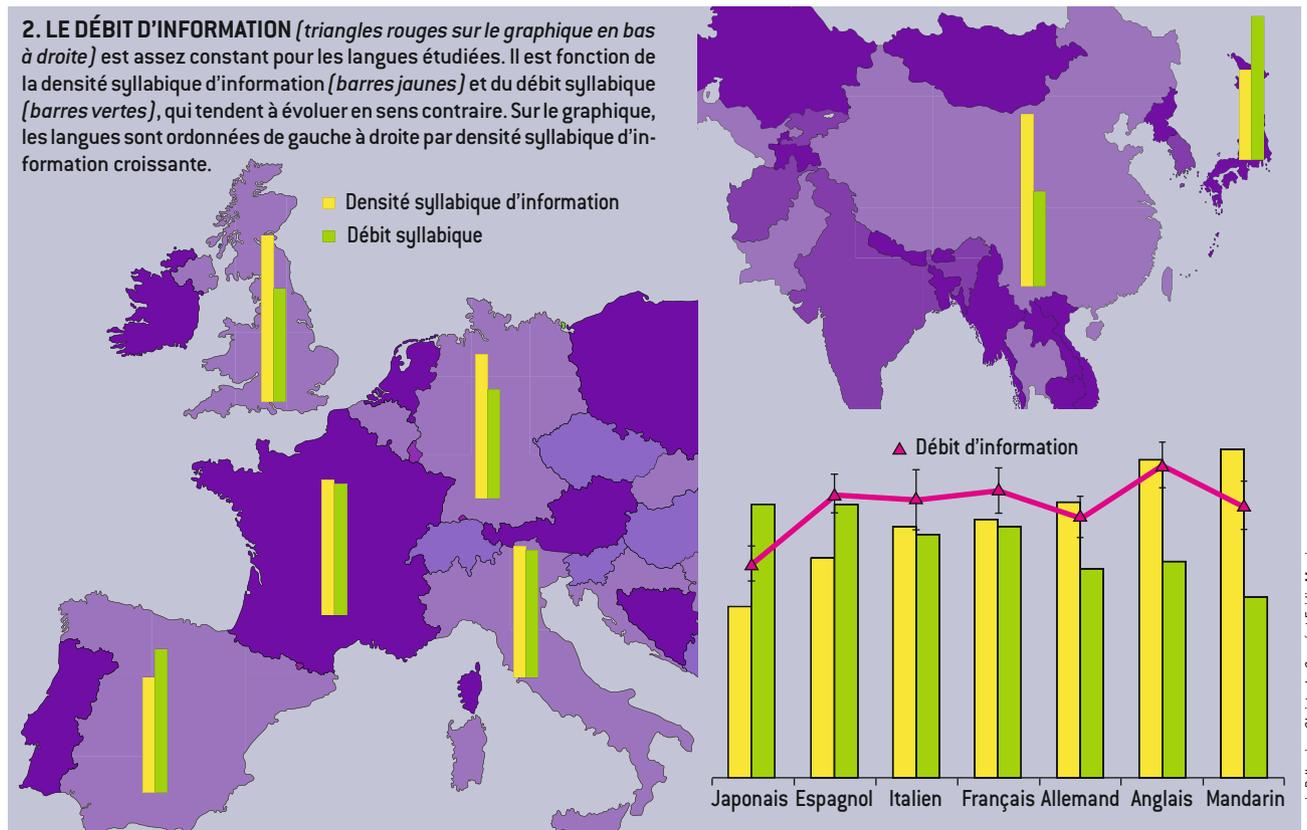
En réalité, d'autres facteurs interviennent, un énoncé n'étant pas constitué de syllabes tirées au hasard. Les syllabes forment des mots : si quelqu'un prononce *an-ti-con-sti-tu-tio-*, la syllabe suivante sera à coup sûr *-nel-* et le fait de l'entendre apporte peu d'information. En outre, les mots 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



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た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.