

[illegible]

MMES

TAGXEDO.COM



A l'occasion du centenaire de la mort de Ferdinand de Saussure, Genève accueillera en juillet le 19^e Congrès international des linguistes. Des origines du langage aux dernières applications concrètes de la discipline, présentation des enjeux de cet événement

Considéré comme le fondateur de la linguistique moderne, Ferdinand de Saussure n'a pas écrit une ligne du «Cours de linguistique générale» qui a fait sa renommée mondiale

La linguistique historique permet d'ouvrir une fenêtre sur l'histoire du peuplement humain et notamment sur l'origine des langues indo-européennes

Dossier réalisé par
Vincent Monnet et Anton Vos



GENÈVE AU SOMMET DE LA LINGUISTIQUE

A l'occasion du centenaire de la mort de Ferdinand de Saussure, Genève accueillera en juillet prochain le 19^e Congrès international des linguistes. Des origines du langage aux dernières applications concrètes de la discipline, présentation des enjeux de cette réunion avec Jacques Moeschler, professeur au Département de linguistique de la Faculté des lettres et président de la Société suisse de linguistique

Dans le cadre du centenaire de la mort de Ferdinand de Saussure, Uni Mail accueillera en juillet prochain le 19^e Congrès international des linguistes. Quel est l'objectif de ce symposium?

JACQUES MOESCHLER: Il faut tout d'abord rappeler que c'est un événement exceptionnel puisque la dernière fois que cette réunion, organisée tous les cinq ans, s'est tenue à

Genève, c'était en 1931 et qu'elle ne repassera probablement plus par ici durant ce siècle. Cette 19^e édition, qui a nécessité trois ans et demi de travaux préparatifs devrait réunir près de 1500 participants autour d'une dizaine de conférenciers

principaux. Elle donnera lieu à près de 800 contributions de très haut niveau qui couvriront tous les domaines de la linguistique actuelle (sociolinguistique, phonologie, morphologie, syntaxe, sémantique, pragmatique, histoire de la linguistique, linguistique informatique, psycholinguistique) à l'exception de la linguistique appliquée qui, pour diverses raisons, fait l'objet de congrès séparés.

Le congrès est placé sous le thème de «l'interface langage-cognition». Pourquoi ce choix?

Le centenaire de la mort de Ferdinand de Saussure, qui est considéré par l'ensemble de

la communauté scientifique comme un des pères fondateurs de la linguistique (lire en pages 18 à 21), offre une opportunité symbolique forte pour l'évaluation des perspectives futures de cette discipline. La recherche actuelle dans tous les champs de la linguistique montre que le lien entre la linguistique et les sciences cognitives va se renforcer de façon significative dans les prochaines décennies. Avec l'émergence de technologies d'imagerie médicale non invasives pour l'observation du cerveau, l'évolution des méthodes statistiques et expérimentales, le développement de la recherche en matière d'acquisition et de pathologies du langage, de nouveaux champs d'investigation se sont récemment ouverts et de nouvelles hypothèses ont été posées. Par ailleurs, les domaines traditionnels de la linguistique – phonologie, morphologie, syntaxe diachronique et synchronique, sémantique et pragmatique – jouent aujourd'hui un rôle croissant dans les recherches cognitives, notamment en ce qui concerne les problèmes liés à l'acquisition du langage.

Où se situe la linguistique genevoise sur la scène internationale?

Le Département de linguistique de l'Université jouit d'une reconnaissance internationale. D'une part, grâce à la place de choix qu'occupe Ferdinand de Saussure dans l'histoire de la discipline. D'autre part, parce que les recherches et l'enseignement qui y sont dispensés aujourd'hui sont d'un très bon niveau. Ce qui explique que nous attirons chaque année une trentaine de nouveaux ►

DR



Jacques Moeschler

«Le lien entre la linguistique et les sciences cognitives va se renforcer de façon significative dans les prochaines décennies»

Le menu du centenaire

► 19 FÉVRIER - 28 MAI

Séminaire : «Reading Saussure and Chomsky», par Claire Forel et Genoveva Puskas, Dpt de langue et littérature anglaises
Uni Bastions, B305
Les mardis, 8h15-10h
contact: claire.forel@unige.ch

► 27 FÉVRIER

Conférence de John E. Joseph (University of Edinburgh)
Institut national genevois
Promenade du Pin 1, 18h30
www.inge.ch

► 28 AVRIL 2013

Conférence de Claire Forel: «Ferdinand de Saussure: un Genevois précurseur de la linguistique moderne»
Lyceum Club international
Promenade du Pin 2, 18h30

► 21 - 27 JUILLET:

19^e Congrès international des linguistes
Uni Mail
www.cil19.org

► 17 SEPTEMBRE:

Conférence d'ouverture du semestre d'automne par Claude Hagège (Collège de France), Uni Dufour U600, 18h30

► 24 - 29 SEPTEMBRE:

«Vénus vagabonde», pièce de Marie Gaulis
Uni Mail MS180, 20h
www.a-c.ch/venusvagabonde

► 30 SEPTEMBRE - 16 DÉCEMBRE:

Ciné-club universitaire: «Mondes parallèles»: Saussure, Flournoy, Helen Smith (Des Indes à la planète Mars), origines et postérité d'un thème
Auditorium Arditi, place du Cirque
les lundis, 20h
www.a-c.ch/ccu

► 20 - 21 NOVEMBRE:

Performances
Poésies en mouvement «Saussure»
soirées verbo-voco-digito-visuelles
Lieu à déterminer, 20h
www.makaronic.ch

Programme complet de toutes les manifestations (régulièrement mis à jour):
www.a-c.ch/saussure

étudiants en formation de base ainsi qu'une dizaine d'autres venus de l'étranger afin de suivre une formation post-grade. Ce chiffre, au vu de l'aspect très théorique et du peu de débouchés qu'offre notre discipline, est assez élevé. Mais ce qui compte surtout, c'est que les étudiants qui suivent ces cours se rendent vite compte qu'on y parle de choses très intéressantes, des choses qu'ils savent parfois implicitement, mais qu'ils ne peuvent expliciter clairement.

Par exemple?

Tout le monde sait que le langage est le propre de l'homme. Cependant, il est plus difficile d'expliquer pourquoi. Surtout depuis que l'on sait que l'être humain partage 98% de son patrimoine génétique avec les primates qui, eux, ne parlent pas. En fait, l'une des conditions qui rend le langage possible est que nous possédons dans notre cerveau des zones spécifiquement destinées à la production et au traitement du langage (aire de Broca et de Wernicke) et qui sont associées à un certain type de connexions neuronales. Or, si ces zones existent également chez les primates, elles sont moins développées et leur cortex moteur, comme chez la plupart des mammifères, n'a que des connexions indirectes avec les neurones du contrôle vocal. Par ailleurs, de récentes recherches en psychologie cognitive ont démontré que ce que l'on appelle «l'attention partagée» est absolument essentielle à l'acquisition du langage.

De quoi s'agit-il?

Pour pouvoir apprendre à parler, un enfant doit pouvoir identifier des séquences de sons et les objets qui leur correspondent avant d'associer ces deux types d'éléments. Or, sans attention partagée, un tel processus n'est pas possible. Une des questions très controversées actuellement est de savoir si les primates possèdent ce qu'on appelle «la théorie de l'esprit», cette faculté cognitive qui nous permet d'attribuer à autrui des états mentaux (intentions, croyances, désirs).

Que sait-on aujourd'hui sur les origines du langage?

Les linguistes ont développé de nombreuses théories sur ce point. En simplifiant à l'extrême, la principale difficulté est d'expli-

quer quand et comment les structures neuronales dédiées au langage ont permis une communication externe et la vocalisation du langage. Même s'il existe encore des incertitudes, nous disposons aujourd'hui de nombreux arguments qui rendent possible l'identification du moment à partir duquel le langage est apparu. Globalement, les hominidés se sont séparés des primates il y a environ 7 millions d'années. «Homo sapiens» est, quant à lui, apparu il y a 200 000 ans environ. Le langage semble être un phénomène plus récent puisque les estimations actuelles penchent pour une fourchette allant de -100 000 ans à -40 000 ans. En s'appuyant notamment sur les dessins rupestres de la Grotte Chauvet (dans le sud de la France), certains chercheurs estiment même que cette apparition est plus récente et que le langage serait apparu il y a 10 000 à 30 000 ans.

Science assez abstraite de prime abord, la linguistique a pourtant de nombreuses applications concrètes. Pouvez-vous en citer quelques-unes?

Dans le domaine de la linguistique computationnelle, de nombreuses équipes, dont celle d'Eric Wherli et de Paola Merlo à Genève (lire en page 22), travaillent sur des analyseurs syntaxiques permettant de produire automatiquement des structures de phrases, ce qui pourrait s'avérer très utile pour la traduction automatique. Un des doctorants du professeur Eric Wehrli, Yves Scherrer, vient d'ailleurs de soutenir une thèse basée sur ce type d'outils décrivant les grandes variations orthographiques et phonétiques entre l'allemand standard et les différents dialectes suisses-allemands. Grâce à un travail cartographique extraordinaire, ce travail montre très clairement le lexique particulier qui est utilisé dans chaque région et la prononciation qui y est associée. Autre exemple: Stéphanie Durrleman, chercheuse post-doc au bénéfice d'un subside du Fonds national de la recherche scientifique, travaille sur le développement de la syntaxe chez les autistes. Dans ce domaine capital, compte tenu des bénéfices que peuvent en attendre les familles concernées, on peut penser que des stratégies de remédiation activées très tôt pourraient permettre de stimuler efficacement l'acquisition du langage.

La linguistique est une science relativement peu connue du grand public. L'année de ce centenaire est-elle une occasion de rectifier un peu son image?

C'est effectivement un enjeu important. En cherchant des subsides pour le Congrès auprès de diverses institutions, je me suis aperçu du déficit d'image dont souffrait la linguistique. Il a en effet fallu beaucoup batailler et argumenter pour obtenir des fonds. Et d'une certaine manière, c'est assez logique.

Pourquoi?

Notre travail n'est pas très visible pour le grand public. La plupart des linguistes sont cantonnés dans le milieu universitaire avec

«Tout un chacun
croit savoir ce qu'est
le langage, d'où il
vient et comment
il fonctionne.
Or, les choses
sont autrement
plus complexes»

un cahier des charges qui touche surtout à la formation intellectuelle de nos étudiants. Par ailleurs, nous sommes encore loin d'être considérés à l'égal des généticiens, des spécialistes de l'imagerie ou des astrophysiciens et nous ne le serons probablement jamais pour la simple et bonne raison que tout un chacun croit savoir intuitivement ce qu'est le langage, d'où il vient et comment il fonctionne. Or, en réalité, les choses sont autrement plus complexes, ne serait-ce que parce que le sens d'un énoncé communique davantage d'éléments que ce que les mots qui le composent signifient. ■

19^e Congrès international des linguistes,
Uni Mail, 22-27 juillet,
www.cihg.org

Publicité

Day one is your time to shine

Day one. It's when you show what you're made of. When the doors are opened and the future lies in front of you. When your views count and making a difference is part of the job. From the day you join us, we're committed to helping you achieve your potential. So, whether your career lies in assurance, tax, transaction, advisory or core business services, shouldn't your day one be at Ernst & Young?

Take charge of your career. Now.
www.ey.com/ch/careers



ERNST & YOUNG
Quality In Everything We Do



L'HOMME QUI AIMAIT LES LANGUES

Mort il y a cent ans, Ferdinand de Saussure a fait entrer la linguistique dans la modernité en développant une méthode scientifique d'investigation du langage. Des travaux dont la sémiotique et le structuralisme sont les héritiers directs.
Portrait

Conformément à la tradition familiale, il aurait dû se faire un prénom dans les sciences de la nature. Arrière-petit-fils d'un naturaliste considéré comme le père de la géologie alpine (Horace-Bénédict), petit neveu d'un des découvreurs des principes de la photosynthèse (Nicolas-Théodore) et fils d'un entomologiste grâce auquel les collections d'insectes du Muséum d'histoire naturelle de Genève comptent parmi les plus riches au monde (Henri), Ferdinand de Saussure a pourtant choisi un autre chemin. Et pas n'importe lequel puisque les trois cours de linguistique générale qu'il a donnés entre 1907 et 1911 à l'Université de Genève (lire en pages 20-21) ont posé les fondements de la linguistique moderne en proposant une méthode scientifique d'investigation du langage.

VOCATION PRÉCOCE

Né le 26 novembre 1857 au sein d'une famille de la grande aristocratie genevoise, Ferdinand de Saussure est l'aîné d'une fratrie de neuf enfants. Excellent élève, il montre très tôt un intérêt prononcé pour l'étude des langues anciennes. A 17 ans, il signe un *Essai pour réduire les mots du grec, du latin et de l'allemand à un petit nombre de racines* dans lequel on devine déjà les prémices de la «méthode saussurienne». L'année suivante, il se lance, en autodidacte, dans l'étude du sanscrit.

Cette vocation précoce aurait toutefois très bien pu ne jamais éclore. Le père de Ferdinand souhaite en effet vivement voir son fils suivre la tradition familiale en embrassant une carrière dans le domaine des sciences naturelles. Bon gré mal gré, Ferdinand obtempère tout en suivant parallèlement les quelques cours de philologie et de linguistique dispensés par

Dates clés

- **26 NOVEMBRE 1857:** naissance de Ferdinand de Saussure à Genève
- **1875-1880:** études de philologie à Leipzig et à Berlin
- **1881-1891:** maître de conférence en grammaire comparée des langues germaniques à l'Ecole pratique des hautes études de Paris
- **1891-1912:** enseigne la grammaire comparée des langues germaniques, du grec et du latin, ainsi que le sanscrit, puis la linguistique générale à Genève
- **1896:** nommé professeur ordinaire à l'Université de Genève
- **1907:** 1^{er} cours de linguistique générale
- **1909:** 2^e cours de linguistique générale
- **1911:** 3^e et dernier cours de linguistique générale
- **22 FÉVRIER 1913:** mort à Vufflens (Vd)
- **1916:** publication du «Cours de linguistique générale»

la Faculté des lettres. L'échec du jeune homme aux examens de chimie qui ponctuent sa première année d'étude (auxquels il ne prend pas la peine de se présenter) aura toutefois raison des ambitions paternelles.

Ferdinand de Saussure a désormais les coudees franches. Il en profite pour gagner Leipzig (qui est alors la «Mecque de la linguistique»). Il restera quatre ans en Allemagne. Le temps de se familiariser avec le vieux perse, le celtique, le slave et le lituanien, de rédiger une thèse et surtout un travail intitulé *Mémoire sur le système primitif des voyelles dans les langues indo-européennes*, qu'il publie en 1878 et qui connaît un succès immédiat.

«Le mémoire de Saussure est intéressant à double titre, explique Jean-Paul Bronckart, professeur honoraire de la Faculté de psychologie et des sciences de l'éducation. D'une part, parce qu'il est aujourd'hui encore considéré par les spécialistes comme un des plus grands et importants ouvrages de linguistique comparée. De l'autre, parce que Saussure y adopte déjà une perspective systémique pour étudier les phénomènes linguistiques et qu'il y avance l'idée absolument nouvelle de l'interdépendance entre le versant phonétique et le versant morphologique de la langue au cours de son évolution. Deux aspects qui seront des idées centrales de ce qui deviendra plus tard sa théorie de linguistique générale.»



La théorie générale que propose Ferdinand de Saussure entre 1907 et 1911 change radicalement l'étude de la linguistique.

Saussure introduit
l'idée que tout
signe linguistique
est composé
du signifiant
et du signifié

Dans l'intervalle, Ferdinand de Saussure gagne Paris en 1880. Outre des cours d'iranien, il fréquente la prestigieuse Ecole des hautes études où il obtient un poste de «maître de conférence de gothique et de vieux allemand». Dix ans plus tard, Saussure, qui jouit alors d'une reconnaissance européenne, est de retour à Genève en tant que professeur extraordinaire «d'histoire et de comparaison des langues indo-européennes». En 1906, il se voit offrir un titre de professeur ordinaire et la chaire de linguistique générale.

C'est dans ce cadre qu'il va délivrer les trois cours qui vont lui valoir une notoriété mondiale. Présentée aux étudiants entre 1907 et 1911, la théorie générale que propose Saussure révolutionne littéralement sa discipline. Jusque-là, la linguistique consistait en effet surtout à retracer l'histoire des langues et à

les comparer entre elles. Avec Saussure, elle change de nature.

Selon le Genevois, il ne faut pas considérer les langues comme des organismes qui auraient une vie et une mort, mais comme des systèmes de signes mis en œuvre par les humains pour communiquer entre eux, au même titre que les signaux maritimes et routiers.

Cette approche, qui annonce les développements que connaîtra la sémiologie à partir des années 1960, innove notamment en distinguant le langage (entendu comme la faculté de pouvoir s'exprimer au moyen de signes), la langue (soit l'ensemble des signes utilisés par une communauté pour communiquer) et la parole (définie comme la mise en œuvre des signes linguistiques dans un contexte précis).

Saussure introduit également l'idée que tout signe linguistique est composé de deux

facettes: le signifiant, qui correspond au son produit pour énoncer un mot, et le signifié, qui renvoie au sens qui y est associé. Il y ajoute le principe de l'arbitraire du signe, selon lequel la relation entre le signifiant et le signifié est arbitraire ou repose sur des conventions.

Ces quelques concepts, qui esquissent à gros traits la pensée saussurienne, ont délimité un champ d'étude que les linguistes n'ont pas fini d'explorer. Tout au long du XX^e siècle, ils ont irrigué les sciences sociales au sens large, notamment au travers de savants comme Claude Lévi-Strauss, Jacques Lacan ou Roland Barthes, qui y ont puisé les fondements du structuralisme. Et il aura fallu l'émergence d'un autre géant, en la personne de Noam Chomsky (lire en page 21), pour que la statue de Commandeur érigée autour du nom de Ferdinand de Saussure après sa mort ne commence à vaciller. ■

À LA RECHERCHE DU COURS PE

Ferdinand de Saussure n'a pas écrit une ligne de l'ouvrage qui a fait sa notoriété mondiale. Son «Cours de linguistique générale» a été publié à titre posthume sur la base de cahiers d'étudiants. Depuis, de nombreux chercheurs s'efforcent de restituer la parole originale du maître

Le paradoxe est de taille. Si le nom de Ferdinand de Saussure est connu dans le monde entier, c'est essentiellement grâce à la publication du cours qu'il a donné à Genève entre 1907 et 1911, qui est considéré comme l'acte fondateur de la linguistique moderne. Or, le savant genevois n'en a pas écrit un mot. Publié à titre posthume en 1916, puis réédité en 1922, le *Cours de linguistique générale* a en effet été rédigé par deux de ses anciens élèves sur la base de quelques cahiers d'étudiants ayant assisté à ses leçons. Depuis, au gré de l'apparition de nouveaux éléments, de nombreux chercheurs se sont efforcés de reconstituer la parole du maître avec plus ou moins de réussite. Si bien qu'aujourd'hui encore, nul ne sait exactement ce qu'a bien pu dire Ferdinand de Saussure, il y

a un peu plus de cent ans, dans l'enceinte feutrée de l'auditoire 105 d'Uni Bastions.

Comme tout génie qui se respecte, Ferdinand de Saussure avait ses petites manies. Soucieux de ne pas perdre ses pensées, il les couchait sur tout ce qui lui tombait sous la main, qu'il s'agisse d'une vieille enveloppe ou du verso d'un faire-part de mariage. Esprit vagabond, il interrompait soudainement certaines phrases pour passer d'une idée à l'autre et cultivait surtout une vraie phobie de l'édition.

Confessant «une horreur malade de la plume» et jugeant que ses réflexions n'étaient pas assez abouties, il a donc très peu publié de son vivant, hormis un mémoire resté célèbre consacré aux voyelles dans les langues indo-européennes. De son cours magistral de lin-

guistique générale, il n'a cependant que laissé des traces éparses disséminées au milieu de ses archives.

BOMBE INTELLECTUELLE

Si la pensée du grand linguiste est parvenue jusqu'à nous, c'est donc en premier lieu grâce aux efforts de Charles Bally et d'Albert Sechey. Conscients d'avoir été les témoins indirects d'un grand moment intellectuel, ces deux anciens élèves de Saussure publient en 1916 une première version du *Cours de linguistique générale* à partir de notes d'étudiants compilées.

«Ce texte, qui fait l'effet d'une véritable bombe intellectuelle dans toute l'Europe, va cependant susciter des réserves croissantes à partir des années 1950, explique Claire Antonella Forel, profes-

Le «Einstein» de la linguistique

Il est à la linguistique ce qu'Einstein est à la physique. Titulaire d'une trentaine de doctorats «honoris causa» et professeur émérite au Massachusetts Institute of Technology (MIT), où il enseigne depuis le milieu des années 1950, Noam Chomsky a, tout comme Saussure l'avait



Noam Chomsky

fait un demi-siècle avant lui, radicalement transformé la manière d'aborder sa discipline.

La «révolution chomskienne» s'ouvre en 1955, lorsque le jeune chercheur (il n'a pas encore 30 ans) présente sa thèse de doctorat. Un travail intitulé «The logical structure of linguistic» qui est

souvent décrit comme la contribution la plus importante dans le domaine de la linguistique théorique du XX^e siècle et qui ne sera publiée dans son intégralité qu'en 1975.

L'approche du langage de Chomsky introduit deux concepts clés : celui de «grammaire

généralisée» et celui de «grammaire universelle». Le premier postule que toutes les langues sont construites à partir d'un nombre de règles et d'éléments fini permettant de produire des phrases en nombre infini. Le second est basé sur l'idée que des principes et des paramètres définissent

l'ensemble des langues humaines possibles. Et selon Chomsky, le véritable objectif de la linguistique consiste précisément à identifier l'architecture de cette «grammaire universelle» afin d'accéder à ce qu'il nomme la «langue interne», à savoir la connaissance qu'un locuteur a de sa langue.

Bien que souvent contestées, les thèses développées par le savant américain ont eu une influence majeure non seulement pour l'émergence des sciences dites cognitives (psychologie, neurosciences, intelligence artificielle...), mais également dans des domaines a priori plus éloignés de la linguistique.

Niels Kaj Jerne, lauréat du Prix Nobel de médecine en 1984, a

ainsi utilisé le modèle génératif de Chomsky pour expliquer le système immunitaire humain, faisant le lien entre structures grammaticales et protéiques.

Scientifique d'envergure mondiale, Noam Chomsky est, par ailleurs, également très connu pour son activisme politique. Opposant de la première heure à la Guerre du Vietnam – il a figuré sur la liste secrète des «opposants politiques» établie sous Nixon –, il est l'auteur de nombreux ouvrages à succès sur la politique étrangère des États-Unis et le fonctionnement des médias. Et c'est sans doute cette double casquette qui lui vaut d'être considéré aujourd'hui comme le «plus grand intellectuel vivant» du monde anglo-saxon.

seure associée au Département de langue et littérature anglaises de la Faculté des lettres et coordinatrice des manifestations du centenaire organisé autour de la mort du savant genevois. Compte tenu du nombre restreint de sources sur lesquelles il s'appuie, il contient en effet un certain nombre d'amalgames et de raccourcis. Il donne également une image beaucoup plus aboutie de la pensée saussurienne qu'elle ne l'est au regard des informations dont on dispose aujourd'hui.»

Une première relecture de l'œuvre de Saussure est entamée à la fin des années 1950, lorsque Robert Godel décide de consacrer sa thèse de doctorat aux sources manuscrites du *Cours de linguistique générale* conservées à la Bibliothèque de Genève.

Présent à la soutenance de ce travail et convaincu par la démarche de Godel, Emile Constantin, un autre ancien élève de Saussure, décide alors de confier ses cahiers au chercheur, apportant au dossier de nouvelles pièces manuscrites auxquelles Bally et Sechehaye n'avaient pas eu accès. Il devient dès lors évident que le texte établi en 1916 demande une approche prudente et a besoin d'être complété par un important appareil critique.

Afin de remettre un peu d'ordre dans les conceptualisations saussuriennes, Tullio De Mauro, qui est alors un des chefs de file de l'école structuraliste italienne, livre en 1967 une édition critique du *Cours de linguistique générale* qui fait encore référence.

RÉSULTAT SPECTACULAIRE

La même année paraît le premier tome (le second sera publié en 1974) du monumental travail entamé par le linguiste alémanique Rudolf Engler. Ce dernier présente le texte publié en 1916 par Bally et Sechehaye fragmenté en 3281 segments en regard desquels figurent les éléments tirés des cahiers d'étudiants leur correspondant ainsi que des fragments de textes de la main de Saussure. Spectaculaire, le résultat obtenu se prête toutefois difficilement à une lecture continue.

Le langage est un système de signes
 1. ~~Il s'agit~~ ^{est un cas particulier de la} ~~théorie des signes~~, ^{mais précède} ~~et pour ce~~
 par ce seul fait, il se trouve déjà dans l'impossibilité absolue d'être une chose simple (ni une chose directement saisissable à notre esprit dans la façon d'être) alors même que dans la théorie générale des signes le cas particulier des signes vocaux n'était ni le plus simple ni le plus complexe de tous les cas particuliers connus; tels que l'écriture, la chiffration etc.
 Mais 2. Ce sera la réaction de l'étude du langage sur la théorie des signes, ^{à jamais} ~~que d'avoir~~ l'horizon nouveau qu'elle aura ouvert, que de lui montrer qu'elle apprendra avoir appris ~~un des~~ ^{des} côtés inhérents du signe, savoir celui d'être transmissible et réel.
 Tout un côté nouveau du signe, à savoir que celui-ci est par sa nature transmissible ne commence à être réellement connu que quand on a vu

Manuscrit de la main de Saussure.

Plus accessibles, les trois ouvrages publiés entre 1993 et 1997 par le chercheur japonais Eisuke Komatsu donnent à lire une partie des trois cours donnés par Saussure dans leur continuité et avec une traduction en anglais. Contestée pour certains problèmes philologiques, cette édition a aussi le défaut d'être dénuée d'appareil critique et d'être basée sur un seul cahier d'étudiant pour chaque cours.

Ce qui a motivé la publication en 2005 par les *Cahiers Ferdinand de Saussure* de l'intégralité des notes prises par Constantin durant le troisième cours ainsi que des fragments de la main de Saussure s'y rapportant.

Intégrant l'ensemble des écrits de Saussure concernant la linguistique générale conservés à la Bibliothèque de Genève ainsi que les manuscrits inédits d'un livre sur la linguistique

générale découverts en 1996 dans l'orangerie de la maison qu'occupait Saussure à Genève (la maison Lullin), les *Ecrits de linguistique générale* (2002) proposent, quant à eux, sous la plume de Simon Bouquet et Rudolf Engler, une nouvelle lecture de la pensée saussurienne en peignant le linguiste genevois «à la fois comme un épistémologue de sa discipline et comme un philosophe, soucieux de dénoncer les illusions de toutes sortes dont le langage est l'occasion pour repenser les fondements de son étude».

Enfin, en février 2010, Daniele Gambarara, président du Cercle Ferdinand de Saussure, a lancé un projet visant à l'édition numérique des manuscrits saussuriens. Une démarche dans laquelle les notes récemment retrouvées par un des neveux du linguiste devraient trouver, elles aussi, un jour leur place. ■



© THINKSTOCKPHOTOS

TRADUCTION AUTOMATIQUE: «NOBODY IS PERFECT»

Traduire un texte à l'aide d'un ordinateur passe, dans l'idéal, par plusieurs étapes: l'analyse lexicale, syntaxique puis sémantique. Le résultat dépend de la performance de ces opérations. Mais le logiciel peut rapidement devenir gourmand en mémoire, en temps et en argent

*«Alice commençait à être très fatigué d'être assis par sa sœur sur la rive, et de n'avoir rien à faire: une fois ou deux, elle avait regarda dans le livre de sa sœur lisait, mais il n'avait pas d'images ni dialogues en elle, et ce qui est l'utilisation d'un livre, pensait Alice, sans images ou de conversation?»**
Ce texte est une traduction du premier paragraphe d'*Alice au pays des merveilles* de Lewis Carroll réalisée par un moteur de recherche célèbre et dominant Internet depuis plus d'une décennie.

Manifestement, le résultat demande encore du travail pour mériter une publication – surtout pour un texte littéraire. En revanche, aussi lacunaire soit-elle, cette traduction a été réalisée dans une soixantaine d'autres langues actuellement à disposition. Ce qui constitue un sérieux avantage et contribue à expliquer pourquoi cet outil est si populaire. «Même les traducteurs professionnels y font appel, ne serait-ce que pour obtenir un premier jet»,

confirme Paola Merlo, professeure associée et responsable du groupe de recherche Computational Learning and Computational Linguistics du Département de linguistique (Faculté des lettres).

Ce système de traduction obtient des scores de fiabilité assez médiocres, situés autour des 80% (selon plusieurs tests réalisés par des traducteurs professionnels comme *redaction.be* ou *www.anyword.fr*), ce qui signifie, en gros, que le lecteur bute sur une erreur lexicale ou

de syntaxe tous les cinq mots. Dans l'exemple ci-dessus, aucune phrase n'est correcte du point de vue de l'orthographe ou de la grammaire. Tout le monde, y compris les développeurs du logiciel eux-mêmes, est d'accord pour dire qu'il doit encore être perfectionné. Cela dit, la remarque est valable pour l'ensemble des produits existant sur le marché. *«Il existe des produits très spécifiques, développés pour de grandes entreprises informatiques, bancaires ou autres, qui font de meilleurs scores mais ils sont plus chers, précise Paola Merlo. En bref, il n'existe toujours pas de logiciel de traduction automatique parfait, même après plus d'un demi-siècle de recherches dans le domaine.»*

La traduction automatique est en effet, avec le calcul des trajectoires balistiques à l'aide du système d'équations différentielles, l'application la plus ancienne de l'informatique. Développée à partir de la fin des années 1940, avec la naissance des premiers ordinateurs, elle était au départ destinée à des fins de renseignements militaires. Les premières langues exploitées par ces programmes étaient d'ailleurs l'anglais et le russe, Guerre froide oblige. *«On peut retracer l'histoire politique et économique du monde à travers les paires de langues utilisées dans la traduction automatique (anglais/russe, anglais/allemand, anglais/arabe, anglais/chinois, etc.)»,* note Paola Merlo.

Ce domaine de recherche fait partie de ce qu'on appelle aujourd'hui la linguistique informatique. Celle-ci désigne en réalité deux branches distinctes. La première prolonge la recherche classique sur les langues en se servant des outils de simulation et de modélisation que fournissent les ordinateurs. La seconde se charge du traitement automatique du langage proprement dit et du développement d'applications. Applications qui ne se bornent d'ailleurs pas à la seule traduction automatique.

LONGUE CHAÎNE DE PROBLÈMES

«On peut citer aussi l'extraction automatique d'informations, souligne la chercheuse genevoise. Devenue très importante depuis l'avènement du Web, cette application cherche à tirer des informations contenues dans un texte et à les organiser afin d'obtenir un résumé, du matériel pour alimenter un éventuel système de questions-réponses, etc. Il existe aussi la gestion des dialogues entre l'humain et la machine, par exemple. Dans ce cas, la chaîne des problèmes à résoudre est longue. Il s'agit de reconnaître la parole (ce qui est un problème acoustique), de la retranscrire en mots, de comprendre ce qu'ils signifient, de chercher les informations pertinentes dans une base de données, de les organiser dans un texte cohérent et de faire prononcer finalement ce dernier par une voix de synthèse.»

Quelle que soit l'application, à partir d'un texte écrit, la première étape à franchir dans le traitement du langage s'appelle «l'affectation des parties du discours». Il s'agit, en d'autres termes, de reconnaître parmi les mots lesquels sont des articles, des verbes, des noms, etc. *«Les logiciels qui se chargent de cette étape sont aujourd'hui très performants, explique Paola Merlo. Ils atteignent des précisions de 98 ou 99% dans de nombreuses langues.»*

Ces scores sont obtenus avec des outils conçus au départ pour n'importe quelle tâche de classification mais qui sont écrits de telle manière qu'ils sont capables de s'adapter à un contexte précis, en l'occurrence une langue. La partie spécifique du programme s'adapte en gardant en mémoire des mots et leurs attributions mais aussi les règles générales d'analyse lexicale.

L'étape suivante est l'analyse syntaxique. La machine doit reconnaître les groupes nominaux, les groupes verbaux, les compléments d'objet direct, etc. C'est une partie ardue qui se joue là et les résultats s'en ressentent. Les meilleurs logiciels, pour l'anglais, parviennent à reconnaître correctement environ 90% des groupes grammaticaux dans un texte. Mais, à un niveau d'évaluation plus élevé, environ la moitié des phrases comportent au moins une erreur de syntaxe.

«Nous sommes très actifs dans ce domaine de recherche, explique Paola Merlo. Notre objectif est d'ailleurs de développer un analyseur syntaxique qui ne se limite pas à une langue mais puisse s'adapter à plusieurs idiomes: français, anglais, espagnol, catalan, japonais, chinois, serbe, tchèque, indi...»

Là aussi, les logiciels sont capables de s'adapter à une langue ou à une autre. Mais ce que ces programmes apprennent, en réa-

lité, ce sont des probabilités d'occurrences de mots, de parties du discours et de leurs relations grammaticales. Cela demande que, pour chaque nouvelle langue, ils s'entraînent en étant confrontés à un grand nombre d'exemples. Ce rôle est dévolu aux corpus.

CENT MILLIONS DE MOTS

Un corpus est une collection de textes qui, dans le meilleur des cas, est un échantillon représentatif d'une langue (articles de journaux, transcription de l'oral, textes littéraires, etc.). Il peut contenir des millions de mots. L'un des plus importants, le British National Corpus, en compte 100 millions. Pour l'analyse syntaxique, ils sont beaucoup plus petits. Ceux qui fournissent les meilleurs résultats sont en plus annotés. C'est-à-dire que des linguistes ont attribué à la main les parties de phrase de tous les textes du corpus.

Dans le cas spécifique de la traduction automatique, les logiciels se basent sur des corpus parallèles, généralement des documents officiels existant en plusieurs langues que l'on trouve notamment dans les actes du Parlement européen.

Après la syntaxe vient la sémantique. Dans ce cas, il s'agit de savoir, par exemple, si le sujet est un agent actif, le verbe une action, et le groupe de mots suivant un complément d'objet direct (*«Jean mange une pomme»*) ou si, au contraire, le sujet est un «expérient» suivi d'un verbe d'état psychologique et de la cause de ce même état psychologique (*«Jean aime Marie»*). Il faut aussi pouvoir distinguer dans la phrase *«J'aimerais réserver un vol de Genève à Paris»* qu'un des deux groupes prépositionnels est le départ et l'autre l'arrivée. Il faut résoudre les formes logiques dès qu'apparaissent des mots comme les quantificateurs (*«combien de pays ont une frontière avec la Suisse?»*), clarifier les anaphores, c'est-à-dire comprendre à quoi font référence les pronoms (*«Jean est à Genève. Je l'ai vu hier»*), etc.

L'outil de traduction automatique testé au début de cet article n'effectue pas (ou seulement en partie) toutes ces analyses. Il se borne à mettre en parallèle les mots ou bouts de phrases en se basant sur un corpus de plusieurs dizaines de milliards de mots issus de livres, de documents de l'ONU ou encore de millions de sites web existant en plusieurs langues. Cela ne suffit pas encore à produire une version lisible du roman de Lewis Carroll. Pas encore. ■

* «Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, and what is the use of a book, thought Alice without pictures or conversation?»

«On peut retracer
l'histoire politique et
économique
du monde à travers
les paires de langues
utilisées dans
la traduction
automatique»

PARLER, PAS SI SIMPLE

Dans un lexique mental qui comprend en moyenne 70 000 entrées, le cerveau humain est capable, grâce à un système de tri d'une formidable efficacité, de reconnaître certains mots avant même qu'ils ne soient entièrement prononcés

C'est simple comme «bonjour», dit-on pour qualifier une action anodine. Pourtant, l'unique fait de prononcer ces deux syllabes et d'être compris par son interlocuteur cache des processus d'une complexité souvent insoupçonnée. Comprendre ce qui se passe durant la production des sons correspondant à un mot et lors de leur interprétation par une autre personne est l'un des objectifs poursuivis par Uli Frauenfelder, professeur de psycholinguistique à la Faculté de psychologie et des sciences de l'éducation. Axés sur l'organisation et le fonctionnement du lexique mental – que l'on pourrait trivialement définir comme un dictionnaire de mots situé quelque part dans notre cerveau –, les travaux de son équipe concernent plus particulièrement les processus de stockage, de production et de réception du langage.

Pour que deux personnes puissent communiquer oralement, il faut, d'une part, que le locuteur soit capable de transformer un concept mental (le sens d'un mot) en une séquence de sons, et, d'autre part, que son interlocuteur soit capable de décoder cette dernière pour en retrouver le sens au sein d'un lexique mental qui, chez un adulte moyen, comprend environ 70 000 mots.

VISION «LOCALISTE»

Globalement, les scientifiques distinguent deux manières de représenter l'information contenue dans le lexique mental. La première est une vision «localiste» selon laquelle les informations relatives à chaque mot sont représentées par un ensemble de neurones localisés dans une zone spécifique du cerveau. Reprenant la distinction saussurienne entre l'image acoustique d'un mot (le signifiant) et

son sens (le signifié), la seconde met en avant une organisation plus diffuse, avec des zones spécifiques associées au traitement phonologique et au traitement sémantique.

Plusieurs résultats expérimentaux tendent à corroborer ce dernier point de vue. En particulier, les travaux menés sur les pathologies du langage indiquent que certains patients atteints de lésions cérébrales spécifiques ont perdu l'accès à certaines catégories de mots uniquement, par exemple les termes associés au mobilier.

«Plus communément, l'expérience du «mot sur le bout de la langue», à laquelle chacun de nous a déjà été confronté, milite aussi pour cette vision des choses, explique Odile Bagou, maître-assistante FNS au sein de la Section de psychologie de la FPSE. Dans ce genre de situation, on est souvent capable de retrouver certaines caractéristiques du mot recherché (par exemple, le premier son, le nombre de syllabes, le genre), mais sans être pour autant capable de le produire précisément. Ce phénomène quasi quotidien illustre bien la fragilité du lien entre sens et son. Il montre aussi que l'on peut parfaitement récupérer le sens du message à transmettre et, en particulier, le sens du mot dont on a besoin, sans pour autant réussir à récupérer les sons qui le composent.»

Quelle que soit la réalité physique du lexique mental, ce qui est acquis aux yeux des scienti-

fiques c'est qu'il ne ressemble en tout cas pas à un dictionnaire dans lequel les mots seraient classés les uns après les autres, de manière linéaire. «Il faut considérer que, pour être interprété par un interlocuteur, chaque phonème est transformé en une sorte de clé, explique Uli Frauenfelder. Cette clé est, en l'espace de quelques centaines de millisecondes, comparée à l'ensemble des mots contenus dans le lexique mental. Sitôt que le terme est identifié, les informations sémantiques et syntaxiques qui lui sont associées deviennent disponibles.»

Les capacités du système ne s'arrêtent pas là. En utilisant des méthodes de *gating*, une technique consistant à présenter successivement des parties de mot de plus en plus longues, il a été démontré que le cerveau humain était capable de reconnaître un mot avant même qu'il ne soit totalement entendu ou écrit.

«L'explication qui semble la plus logique est que les mots sont organisés en cohorte, précise Uli Frauenfelder. Les termes qui partagent certains éléments, comme un premier son identique, sont ainsi regroupés au sein du lexique.

Sachant qu'il existe près de 2000 mots commençant par le son /r/, mais qu'ils ne sont plus qu'une dizaine lorsqu'on y ajoute le son /i/, c'est un moyen extrêmement efficace pour identifier rapidement la bonne cible.»

Cette faculté est d'autant plus nécessaire que l'estimation de la taille du lexique mental ne considère en général qu'une seule prononciation, dite «canonique», par mot. Or dans





IMAGES: ISTOCK

les faits, il est fréquent qu'un mot se prononce de plusieurs manières. Ce phénomène est flagrant en anglais, où «international» se réduit parfois à «i'national», mais également en français, les Genevois étant souvent désignés comme des «G'nevois» ou des «Gen'vois».

Sur ce point, de récentes recherches menées conjointement par Uli Frauenfelder et Audrey Bürki, maître assistante au sein de la Section de psychologie, ont permis de montrer que les variantes de prononciation étaient représentées en mémoire au même titre que les prononciations canoniques.

«PETIT AMI» OU «PETIT TAMIS»?

«Ces travaux remettent en question les modèles traditionnels du système de production langagier, commente Uli Frauenfelder. Ils suggèrent que le système de stockage et de production des mots repose sur des principes bien plus complexes qu'on ne l'imaginait jusqu'ici.»

Par ailleurs, comment expliquer que l'auditeur puisse accéder correctement au mot «ami» lorsqu'il entend «petit ami», alors que cette séquence de sons pourrait également correspondre à la séquence «petit tamis»? Contrairement au langage écrit, la parole est en effet continue et ne comporte ni espace ni silence permettant à l'auditeur de localiser où et quand débute ou finit un son, une syllabe ou un mot.

Les recherches menées conjointement par Uli Frauenfelder et Odile Bagou

visent précisément à étudier le rôle des différents indices disponibles dans le signal de parole et permettant de le découper correctement en mots.

«Les études antérieures ont montré que plusieurs indices pouvaient aider l'auditeur, poursuit le professeur. Par exemple, on sait que les sons qui se situent au début des mots n'ont pas les mêmes caractéristiques acoustiques que ces mêmes sons dans une autre position. On sait également que nos connaissances de la langue, et en particulier, les connaissances phonologiques, sont utilisées efficacement pour extraire les mots de la parole continue. Par exemple, le fait que le son /n/ ne puisse pas précéder le son /l/ en début de mot en français nous indique la présence d'une frontière entre les mots «bonne» et «lutte» dans la séquence «bonne lutte». L'accentuation et la mélodie de la parole (la prosodie) peuvent aussi nous aider à déterminer où se situent les fins de mots en français puisque les mots sont usuellement accentués et portent des variations mélodiques sur leur dernière syllabe. Ce qu'on ignore en revanche, c'est la manière dont ces différents indices collaborent ou deviennent antagonistes.»

Les résultats obtenus récemment par les deux chercheurs montrent que ces indices sont hiérarchiquement organisés, selon un ordre qui dépend de la langue du locuteur. En anglais, les indices prosodiques sont ainsi moins pertinents que les indices phonologiques, tandis que la situation est inversée en français. Ce qui indique que la hiérarchie des différents indices permettant de segmenter la parole ne serait donc pas universelle, mais spécifique à la langue du locuteur. ■

«Chaque phonème est transformé en une sorte de clé qui, en quelques millisecondes, est comparée à l'ensemble des mots contenus dans le lexique mental»



Paysan d'Anatolie, lieu où aurait été parlé, il y a une dizaine de millénaires, l'ancêtre commun des langues indo-européennes avant qu'il ne se répande à travers Europe et vers l'Inde avec les premiers agriculteurs. La langue turque, parlée aujourd'hui dans cette région, ne fait pas partie de cette famille linguistique, contrairement au kurde.

LA LONGUE MARCHÉ DES LANGUES

Une étude basée sur la linguistique place l'origine des langues indo-européennes en Anatolie d'où elles se seraient diffusées, à partir de 8500 av. J.-C., vers l'est jusqu'en Inde et vers l'ouest en Europe. Elle contredit la thèse, jusqu'ici dominante, d'une naissance dans les steppes pontiques vers 5000 av. J.-C.

Il existe une controverse tenace sur l'origine des langues indo-européennes. Cette famille, qui regroupe entre autres les idiomes celtiques, italiques, germaniques, balto-slaves, indo-iraniens ou encore grecs, serait issue, pour une partie des linguistes, des nomades des steppes pontiques (au nord de la mer Noire, entre le Dniepr et la Volga) qui auraient déferlé au cours de l'âge du Bronze (vers 4000 av. J.-C.) sur le Vieux Continent avec leurs chars à deux roues tirés par des chevaux. Cette thèse est soutenue par un certain nombre d'éléments linguistiques et archéologiques. Pour les autres, elle aurait été diffusée quelques millénaires plus tôt par les premiers paysans venus d'Anatolie, en Turquie actuelle, qui ont fait basculer progressivement l'Europe dans le Néolithique.

Sans trancher définitivement la question – une dispute scientifique ne s'éteint pas si facilement –, un article paru dans la revue *Science* du 24 août 2012 fait franchement pencher la balance du côté de la seconde option. Les auteurs ont fait appel à la puissance des ordinateurs pour simuler la diffusion des langues dans le temps et l'espace. Résultat : les langues indo-européennes viennent d'une langue originelle parlée il y a 8500 ans dans le sud de la Turquie. Russel Gray, linguiste, professeur à l'Université d'Auckland en Nouvelle-Zélande et l'un des auteurs de l'étude, a détaillé la démarche à l'occasion du Congrès sur la diversité génétique et l'origine des Européens qui s'est tenu à Genève les 15 et 16 janvier dernier. Un événement organisé par Alicia Sanchez-Mazas, professeure à la Faculté des sciences

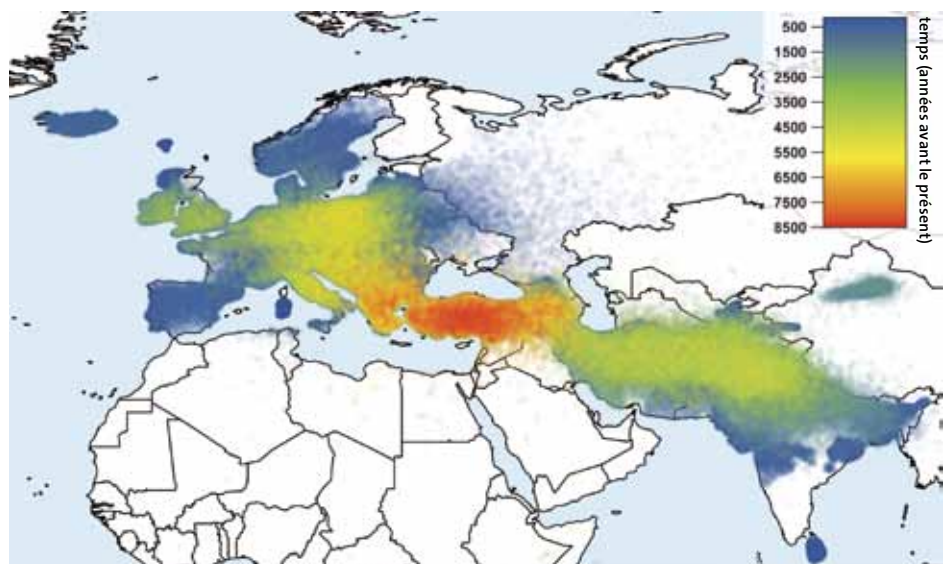
et directrice du Laboratoire d'anthropologie, génétique et peuplements.

PERTES ET GAINS

«Nous avons utilisé un modèle développé initialement pour rechercher, à partir de données moléculaires, l'origine des épidémies virales, explique Russel Gray. Nous l'avons adapté pour analyser un ensemble de mots appartenant au vocabulaire de base de 103 langues indo-européennes contemporaines et antiques. A la place des mutations génétiques qui sont la cause de l'évolution en biologie, nous avons utilisé les pertes et les gains au cours du temps de «cognats», ou mots apparentés entre les langues.»

Pour être enregistrés comme tel, les mots doivent avoir une signification similaire et, plus important encore, montrer des correspondances systématiques dans leur ►

«Nous avons utilisé un modèle développé initialement pour rechercher, à partir de données moléculaires, l'origine des épidémies virales»



Reconstruction spatiale de l'expansion des langues indo-européennes basée sur un échantillon de 103 idiomes.

prononciation. Par exemple, le mot anglais pour cinq, *five*, possède des mots apparentés en allemand (*fünf*), en suédois (*fem*) et néerlandais (*viijf*), reflétant une descendance du mot proto-germanique (**fimf*). Des mots apparentés peuvent disparaître ou apparaître, par exemple lorsqu'une langue commence à utiliser un nouveau terme pour une signification donnée. Ainsi, dans la lignée qui a mené vers les langues germaniques, le mot indo-européen pour os, **ast(h)*, a été remplacé à un certain moment par **bainan*, ce qui a donné *bone* en anglais, *Bein* en allemand, *benen* en suédois ou encore *been* en néerlandais.

A ces «mutations» linguistiques, les chercheurs ont ajouté les distributions géographiques de ces langues et des contraintes de diffusion avant de faire remonter le temps à leur simulation. L'exercice a été répété plusieurs fois en variant un certain nombre de paramètres.

Dans tous les cas de figure, les ordinateurs ont désigné l'Anatolie comme point de départ des langues indo-européennes. Le résultat est encore plus significatif si l'on ne prend en compte que la vingtaine de langues aujourd'hui éteintes (hittite, gothique, latin...). Et il pointe toujours la Turquie si le modèle ne considère que les langues vivantes.

TOUTES LES LANGUES MÈNENT À ROME

Sachant qu'historiquement toutes les langues romanes sont issues du latin, les chercheurs ont choisi ce sous-groupe pour vérifier leur modèle. Celui-ci n'a pas failli et a calculé que l'endroit où la probabilité est la plus grande de trouver la langue originelle est la région de Rome.

«Les premières phases du scénario que nous proposons sont cohérentes avec l'expansion de l'agriculture qui traverse l'Europe via les Balkans et atteint les côtes occidentales vers 5000 av. J.-C., poursuit Russel Gray. Il correspond également avec les résultats génétiques et craniométriques qui montrent qu'il existe une contribution anatolienne dans le génome européen.»

Russel Gray n'en est pas à son coup d'essai. Il a mené, il y a quelques années déjà, une étude similaire sur les langues austronésiennes, parue dans la revue *Science* du 23 janvier 2009. Cette famille de langues comprend quelque 1200 idiomes parlés dans presque tout l'océan Pacifique. Il existe un accord assez généra-

Les linguistes sur les traces des migrants

Plusieurs linguistes étaient présents lors du Congrès sur la diversité génétique et l'origine des Européens qui s'est tenu à Genève les 15 et 16 janvier derniers. L'un d'eux, Laurent Sagart, du Centre de recherches linguistiques sur l'Asie Orientale à Paris, était l'un des modérateurs. Il explique les raisons de sa présence dans une assemblée dominée par les archéologues et les biologistes.

Campus: Que font des linguistes dans un congrès sur la migration et le peuplement?

LAURENT SAGART: Dans ma discipline, il existe une catégorie de chercheurs, les linguistes historiques dont je fais partie, qui étudient l'origine des langues. L'histoire du peuplement humain nous intéresse dans la mesure où les populations emportent avec elles les langues qu'elles parlent lorsqu'elles se déplacent. L'une de nos tâches consiste à reconstruire des langues disparues. Du coup, nous cherchons aussi à reconstruire leur histoire, comprendre qui les parlaient, à quel endroit et à quelle époque. Dans cette optique, la linguistique est une des fenêtres sur le passé, au même titre que l'archéologie et la génétique des populations.

La linguistique est-elle aussi fiable que l'archéologie ou la génétique dans ce domaine?

Elle est fiable. Nous disposons de méthodes permettant d'exclure le hasard. Quand on veut reconstruire une langue morte, l'ancêtre

des langues indo-européennes par exemple, nous commençons par repérer des mots qui se ressemblent d'une langue à l'autre. Ils sont choisis dans une partie du vocabulaire qui est connue pour résister aux emprunts. Il faut que ces mots soient liés entre eux en fonction d'un système de règles de correspondances. Ces lois, bien établies, permettent d'éviter les mots qui se ressemblent par hasard (il y en a beaucoup) ainsi que les mots em-

pruntés d'une langue à l'autre. En agissant ainsi, on peut tenter de reconstruire une partie du vocabulaire de la langue ancestrale et avoir une idée de la manière dont ces mots se prononçaient et ce qu'ils voulaient dire.



Laurent Sagart

Les langues ouvrent-elles aussi une porte sur la culture du passé? Une partie du vocabulaire que nous obtenons se retrouve dans

toutes les langues. Elle est universelle et ne nous dit rien sur la culture des populations. Mais une autre partie, elle, véhicule une information utilisable. On peut déterminer si une langue ancestrale possédait des mots pour désigner les céréales, par exemple, ou pour l'action de récolter. Ces termes indiquent si les locuteurs en question pratiquaient une certaine forme d'agriculture ou si, au contraire, on a affaire à une culture pastorale. Il est ensuite possible de corréliser ces résultats avec des données archéologiques qui peuvent fournir des dates absolues.

pour affirmer que l'ancêtre commun de toutes ces langues était parlé à Taïwan il y a plusieurs millénaires. La simulation élaborée par Russel Gray et ses collègues a confirmé cette thèse et a permis de montrer que la diffusion des langues austronésiennes à travers les îles aurait débuté il y a 5200 ans et serait passée par des périodes de pause et d'accélération liées à des innovations techniques et sociales.

Les critiques de cette méthode numérique ne manquent pas, cependant. Les tenants de la thèse des steppes pontiques notent, par exemple, que les chercheurs ont limité leur étude au vocabulaire qui ne représente qu'une

partie minoritaire des données accessibles en linguistique. Ils estiment également que trop de paramètres concernant la diffusion et le taux de modification des langues sont arbitraires.

En réponse, les auteurs de l'étude sur les langues indo-européennes précisent que les archéologues, eux, doivent lutter pour lier leurs découvertes à des cultures particulières. *«Guigner dans le passé de l'être humain n'est pas une chose simple, expliquent-ils. Cela revient à tenir une bougie au-dessus d'un sombre abysse. On doit utiliser la moindre information à disposition.»* ■

GÈNES ET MOTS EN ASIE DE L'EST

Une étude génétique et linguistique soutient que l'Asie de l'Est a été peuplée via deux routes contournant l'Himalaya par le sud et par le nord

Le peuplement de l'Asie de l'Est par l'homme moderne est un processus très complexe pour la compréhension duquel la linguistique, une fois de plus, peut apporter une aide précieuse. Selon l'une des théories dominantes, la migration se serait opérée par une route venant d'Afrique et passant par le sud de l'Himalaya. Cette population originelle aurait ensuite essaimé vers le nord et aurait donné naissance à tous les habitants de l'Asie de l'Est dont la plus grande partie est occupée par la Chine actuelle.

MOUVEMENT VENU DU SUD

Dans un article paru le 4 janvier 2011 dans la version en ligne de la revue *American Journal of Physical Anthropology*, Da Di et Alicia Sanchez-Mazas, respectivement assistant et professeure au Laboratoire d'anthropologie, génétique et peuplements de la Faculté des sciences, remettent en question cette vision. Leurs principaux résultats, basés sur des données génétiques provenant d'études en partie chinoises qui, n'ayant jamais été traduites, sont restées ignorées jusque-là en Occident, soutiennent un modèle alternatif, dit «de la pince». Il en ressort que le peuplement de l'Asie de l'Est aurait commencé, il y a entre 50 000 et 60 000 ans, par un mouvement venu du sud mais aurait été suivi quelques dizaines de millénaires plus tard par une migration ayant contourné l'Himalaya par le nord. Les données révèlent même que les groupes de migrants du sud étaient relativement isolés et de petite taille, contrairement à ceux du nord qui auraient vécu une histoire démographique beaucoup plus stable. Les deux courants se seraient rencontrés quelque part au nord du fleuve Bleu (Yangzi Jiang) avant que les nouveaux arrivants ne se mélangent progressivement avec les populations autochtones au cours de leur marche vers le sud.

Dans le cadre de ce travail, la linguistique a d'abord servi à cataloguer les populations actuelles de l'Asie de l'Est selon leur langue. Ces idiomes sont classés en cinq familles. Les

langues altaïques (mongoles, tounghouses; turques), sino-tibétaines (chinoises et tibéto-birmanes), hmong-mien (parlées au sud de la Chine, nord du Laos et du Vietnam), tai-kadai (dont fait partie le thaïlandais) et austro-asiatiques (vietnamien, khmer, etc.).

teurs des langues altaïques, les Japonais et les Coréens. Cette observation tendrait donc à soutenir une hypothèse linguistique – longtemps repoussée mais actuellement de moins en moins contestée – selon laquelle toutes ces langues feraient partie de la même famille, elle-

IMAGE FORUM



Jeune fille de l'ethnie Hmong, au nord du Vietnam.

«Entre la génétique et la linguistique, on ne travaille pas sur la même échelle de temps, précise Da Di. Si la première permet de remonter dans l'histoire du peuplement humain sur des dizaines de milliers d'années, la seconde, qui cherche à reconstruire les ancêtres communs des langues actuelles, se limite aux dix derniers millénaires, au grand maximum, du moins selon l'avis de la majorité des linguistes.»

Cela dit, dans la plupart des cas, les populations parlant des langues apparentées présentent également moins de différences génétiques. Ainsi, l'étude genevoise a montré qu'existe une affinité génétique entre les locu-

même étant liée, selon de nombreuses études, à des populations migrant en Asie de l'Est par une route passant par le nord de l'Himalaya.

Il semble par ailleurs que les Sino-Tibétains du nord soient plus proches, du point de vue génétique, des Altaïques, tandis que les Sino-Tibétains du sud s'apparenteraient davantage aux Hmong-mien, Tai-kadai et Austro-Asiatiques. Ce profil génétique est cohérent avec la distribution linguistique puisque le mandarin, qui est parlé au nord, a été influencé par les langues altaïques. Au sud, on trouve un assemblage d'un grand nombre de langues ou dialectes chinois (dont le cantonais). ■