

Moisson de jupiters

Trois planètes de la taille de Jupiter hors du Système solaire.

Les nuits d'été sont propices à l'admiration de la voûte céleste. Les observateurs attentifs retracent les constellations : vers minuit, à la mi-juillet, la Grande Ourse est au Nord-Ouest, la Vierge se couche à l'Ouest, Pégase se lève vers l'Est. Ces régions du ciel méritent notre attention : on y a découvert les premières planètes extra-solaires en orbite autour d'étoiles analogues à la nôtre.

Une étoile entourée de planètes tourne autour du centre de masse du système : pour un observateur terrestre, elle s'approche et s'éloigne périodiquement. Lorsqu'elle s'éloigne, les raies de son spectre lumineux sont décalées vers les grandes longueurs d'onde, par effet Doppler (comme la sirène d'une ambulance qui s'éloigne semble plus grave) ; lorsqu'elle se rapproche, les raies sont décalées vers les courtes longueurs d'onde. Ces variations de longueur d'onde des raies permettent de calculer la vitesse de l'étoile sur son orbite. Comme on connaît la masse de l'étoile, qui dépend de sa luminosité, les lois de Kepler indiquent, sans équivoque, les masses des planètes et leur distance à l'étoile.

Par cette méthode, trois corps ont été découverts en orbite autour de trois

étoiles de même masse que le Soleil : le premier, de la masse de Jupiter, à 0,05 unité astronomique de 51 de Pégase (l'unité astronomique est la distance moyenne de la Terre au Soleil), le second, de 2,8 fois la masse de Jupiter, à 2 unités astronomiques de 47 de la Grande Ourse et le troisième, de 6,4 fois la masse de Jupiter, à 0,5 unité astronomique de 70 de la Vierge. Dans le Système solaire, Jupiter est à 5 unités astronomiques du Soleil.

Quelle est la nature de ces corps ? Les planètes du Système solaire sont issues de l'agrégation de poussières en orbite au sein d'un disque protoplanétaire. Même les planètes géantes gazeuses ont un noyau solide. Le noyau de Jupiter, environ deux fois plus gros que la Terre, a attiré, par gravitation, l'hydrogène qui compose aujourd'hui 90 pour cent de la planète.

Une planète peut aussi se former par effondrement d'un nuage de gaz, comme une étoile. L'évolution d'un nuage de gaz qui s'effondre dépend de sa masse. Si cette masse dépasse 80 fois celle de Jupiter, les gaz comprimés au centre sont suffisamment chauffés pour que la réaction de fusion de l'hydrogène en hélium démarre : une

étoile analogue au Soleil se forme. Si la masse du nuage est comprise entre 13 et 80 fois celle de Jupiter, seule la réaction de destruction du deutérium en protons et en neutrons s'enclenche : l'étoile reste une naine brune. Enfin, si la masse est inférieure à 13 fois celle de Jupiter, aucune réaction n'est possible : la boule de gaz est alors nommée planète. Lors de la formation de l'étoile, le nuage de gaz peut se couper en deux, et donner naissance à un système d'étoiles doubles, ou à une étoile accompagnée d'une naine brune ou d'une planète.

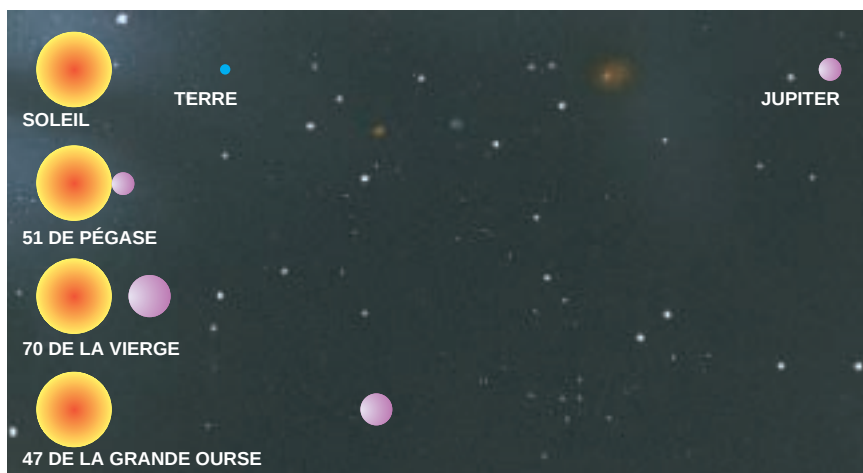
GAZ OU POUSSIÈRES ?

Les paramètres orbitaux mesurés par les astronomes permettent parfois de trancher entre les deux scénarios de formation. L'orbite de la planète qui gravite autour de 70 de la Vierge est une ellipse très excentrée. La formation à partir d'un disque protoplanétaire est donc peu probable : au sein d'un tel disque, les collisions aléatoires que subissent les poussières situées sur des orbites elliptiques transforment celles-ci en orbites circulaires (notons que l'orbite de Mercure, qui n'est pas circulaire, est une exception expliquée par l'attraction gravitationnelle des autres planètes du Système solaire). Si d'autres planètes perturbaient l'orbite de celle observée autour de 70 de la Vierge, elles seraient très massives et on les aurait observées aussi.

La découverte de ces premières planètes, et d'autres qui suivront bientôt, permettra de mieux comprendre la formation et la structure du Système solaire. On n'avait jusqu'à présent aucun point de comparaison sûr : la présence de planètes n'est que soupçonnée autour de β -Pictoris, une étoile semblable au Soleil, et l'histoire du système planétaire en orbite autour du pulsar 1257+12, une étoile à neutrons, est très différente de la nôtre.

En outre, si les planètes de 51 de Pégase, 70 de la Vierge ou 57 de la Grande Ourse se sont formées à partir d'un disque protoplanétaire, on peut spéculer sur l'existence de satellites de la taille de la Terre, constitués majoritairement de roches. Ces satellites pourraient-ils abriter la vie ? La température de surface du compagnon de 70 de la Vierge serait de 80 degrés, compatible avec la présence d'eau liquide. Les astronomes se mobilisent pour explorer plus avant ces nouveaux mondes. Le CNRS a décidé de constituer un regroupement national sur ce thème.

Jean SCHNEIDER
CNRS, Observatoire de Paris



Pour la première fois, des planètes ont été découvertes autour d'étoiles analogues au Soleil : 51 de Pégase, 70 de la Vierge et 47 de la Grande Ourse. La masse de ces planètes est comprise entre 1 et 6,4 fois la masse de Jupiter (les étoiles ne sont pas à l'échelle).

Le sens des mots

La statistique de la distance entre deux apparitions du même mot permet de distinguer deux catégories de mots.

Tous les mots de cette phrase n'ont pas la même importance pour sa compréhension. Pour s'en persuader, il suffit de la relire en supprimant soit les huit mots *Tous, les, de, cette, n, ont, la, sa*, soit un seul des trois mots *phrase, importance* ou *compréhension*. Dans le premier cas, la phrase est encore compréhensible, dans le second, plus du tout. Certains mots organisent la syntaxe de la phrase : ce sont les mots que nous nommons «mots grammaticaux». D'autres servent plutôt à exprimer le sens, ce sont les «mots pleins». Nous présentons ici une méthode permettant de distinguer ces populations de mots selon la distance en mots séparant deux occurrences d'un même mot. La base de données utilisée est composée de deux années d'articles du quotidien *Le Monde*, soit environ 40 millions de mots.

Imaginons un texte où un mot *M* est réparti au hasard : en chaque position, la probabilité de trouver le mot *M* est indépendante de la nature des mots qui le précèdent. La loi de probabilité de la distance séparant deux apparitions de *M* est appelée, dans ce qui suit, loi de distribution aléatoire. Réciproquement, si la loi de distribution d'un mot est la loi de distribution aléatoire, la probabilité de le rencontrer en une position ne dépend pas des mots qui l'entourent.

En examinant la loi de distribution du mot *de*, on constate qu'excepté pour des petites distances (jusqu'à six mots), cette loi est très proche de la loi aléatoire : on rencontre le mot *de* dans un texte comme s'il y était réparti aléatoirement. En se fixant un seuil arbitraire de «ressemblance» par rapport à une distribution totalement aléatoire, nous pouvons établir une liste de mots qui ont ce comportement : par exemple, les prépositions, *de, par*, les pronoms, *en, il*, les déterminants, *la, les*, les formes conjuguées des verbes *mettre, faire*, qui sont des verbes support de la langue (verbes auxiliaires utilisés dans la nominalisation), les adverbes *probablement* et

relativement. Autrement dit, les mots grammaticaux ont une distribution quasi aléatoire. Le choix du seuil de vraisemblance est arbitraire, mais un mot ayant généralement plusieurs significations, la loi calculée est une moyenne sur ses différentes significations.

Pour les mots que nous ne savons classer dans l'une ou l'autre des catégories, la méthode fera un choix qui dépendra du seuil. Plus la tolérance sera faible, plus les mots trouvés seront «purement grammaticaux». Par exemple, le verbe *mettre* est un mot plein dans les phrases où il exprime le déplacement d'un objet, mais c'est un mot grammatical lorsqu'il est utilisé comme verbe support (comme dans *mettre dans l'embarras*).

Ce critère, qui permet de distinguer les mots grammaticaux n'est pas un critère de fréquence : certains termes fréquents comme *France* n'ont pas une distribution aléatoire. L'écart de ces distributions à la loi aléatoire pour les petites distances est de nature syntaxique : trouver deux occurrences successives du mot *de* est très peu probable, alors que, pour une distribution aléatoire, cette probabilité serait supérieure à celle de trouver deux occurrences séparées par un autre mot. De surcroît, la signature d'un mot (c'est-à-dire l'allure de la loi de probabilité) est relativement stable par

changement de corpus : dans des corpus aussi variés qu'une traduction de la Bible et un quotidien, la signature du mot *de* est semblable. La signature des mots grammaticaux ne dépend pas du contenu sémantique des textes, mais du style de l'auteur.

LES MOTS PLEINS

En revanche, on constate que la distribution des mots pleins (qui intuitivement semblent être «porteurs de sens») n'est pas aléatoire : la présence de certains mots dans une partie du texte dépend des autres mots de cette partie. Cette dépendance entre les différents termes est associée au contexte. Dans un article traitant de politique, il est plus probable de trouver des mots comme *ministre* ou *gouvernement*, que dans un article traitant de sport. Réciproquement, un article contenant le mot *gouvernement* traite plus probablement de politique que de sport. Ces deux remarques entraînent que, dans un article contenant le mot *gouvernement*, il est plus probable qu'ailleurs de trouver le mot *ministre*. On dit alors que le mot *gouvernement* a un effet de contexte positif sur le mot *ministre*, et le mot *match* a un effet de contexte négatif sur le mot *ministre*. Nous avons quantifié cet effet de contexte, en comparant la loi de distribution d'un mot dans un contexte donné, avec la loi de distribution aléatoire qu'il aurait s'il était à comportement grammatical.

Les perspectives ouvertes par ces études sont au moins triples. La première est l'étude de méthodes de compression de texte (ou d'index) performantes, fondées sur la finesse de la description de la loi de distribution des mots. La deuxième est l'élaboration de catégories sémantiques autour d'un mot, par l'examen des effets de contexte. À terme, on pourrait lever l'ambiguïté sur le sens des mots en étudiant leurs contextes, ce qui pourrait être appliqué à la traduction automatique. Par exemple, pour le mot *jeu*, nous avons, d'une manière totalement automatisée, obtenu trois ensembles de mots appartenant à trois champs sémantiques distincts : le contexte sportif, le contexte théâtral, le contexte musical et nous avons aussi obtenu une liste d'expressions contenant le mot *jeu* employé dans n'importe quel contexte. Enfin, on pourrait caractériser un auteur par les signatures, dans son œuvre, des mots à comportements grammaticaux.

Jean SENELLART – LADL, Paris

	PC	PS	RPR	UDF	écologie	spectacle
Barre	0,65	1,29	3,78	4,13	1,89	-1,55
Chirac	-0,08	0,86	5,97	3,64	1,49	-0,65
Cresson	-0,37	1,49	0,32	0,19	1,12	-1,30
Lalonde	0,28	2,77	1,19	1,66	9,43	-0,19
Marchais	2,51	2,09	2,34	1,23	5,02	-4,32
Mitterrand	-1,18	0,64	1,09	0,73	0,56	-1,22

Ces valeurs mesurent la probabilité de trouver une mention d'un personnage politique (colonne) dans un bloc de 100 mots autour de la mention d'un parti politique ou d'un autre mot (ligne). Plus la valeur est élevée, plus l'effet de contexte est important. La valeur maximale pour un personnage politique s'obtient généralement pour son propre parti, sauf pour le mot *Mitterrand* qui présente un comportement plus grammatical. Le mot *spectacle* à un effet de contexte négatif pour toutes les mentions de personnages politiques.

Les membranes de nos ancêtres

On trouve dans le pétrole la composition chimique des premières cellules vivantes.

Comment les premières cellules vivantes se sont-elles formées? La réponse que proposent Guy Ourisson, Yoichi Nakatani et leurs collaborateurs du Centre de neurochimie de Strasbourg se fonde sur des études *a priori* sans relation avec l'étude de la chimie prébiotique : l'analyse chimique du pétrole.

Il y a une trentaine d'années, G. Ourisson et Pierre Albrecht étudiaient les molécules constituant la matière organique des sédiments : pétroles, charbons, argiles, schistes, calcaires, etc. Ils observaient que la présence d'une molécule nommée cholestane, dans de nombreux sédiments, est un indice de la présence initiale de cholestérol (ce dernier, que l'on considère comme un ennemi de nos artères, est un constituant des membranes cellulaires de tous les êtres vivants, à l'exception des bactéries). Le cholestane est fréquemment accompagné de ses dérivés portant un ou deux atomes de carbone supplémentaires, structures que l'on retrouve dans les stérols des plantes ou de certains micro-organismes. Les «fossiles moléculaires» de stérols sont souvent accompagnés de produits nommés diastéranes, dont on peut obtenir le squelette par chauffage modéré des dérivés du cholestane avec des argiles : cette «simulation» renforce l'hypothèse d'une présence de cholestérol d'êtres vivants ayant conduit, par chauffage sous pression, dans les entrailles de la Terre, à la formation des pétroles.

Lors de cette étude de la matière organique des sédiments, les chimistes trouvèrent des composés qui furent nommés «géohopanoïdes», dérivés d'une molécule nommée hopane, à 30 atomes de carbone. Cette molécule est un triterpène, c'est-à-dire qu'elle est composée par l'enchaînement de motifs isoprène (voir la

figure), dont le squelette est formé de cinq atomes de carbone.

Les terpènes sont des molécules abondantes dans le monde naturel : on les trouve dans l'essence de térébenthine, dans l'écorce des bouleaux ou dans le parfum des roses ; ce sont des hormones de plantes, des constituants de membranes cellulaires d'archéobactéries, des hormones sexuelles d'animaux, des pigments de légumes... Leur structure fortement ramifiée rend leur biodégradation difficile, ce qui explique leur pérennité dans les sédiments.

Quelle était l'origine de ces géohopanoïdes? Les chimistes étaient perplexes, car aucun produit naturel connu n'avait une structure moléculaire ayant pu, même par maturation dans les sédiments, se transformer en une molécule ayant cinq atomes de carbone de plus que le hopane,

au demeurant rare. Les géohopanoïdes étaient des structures «orphelines».

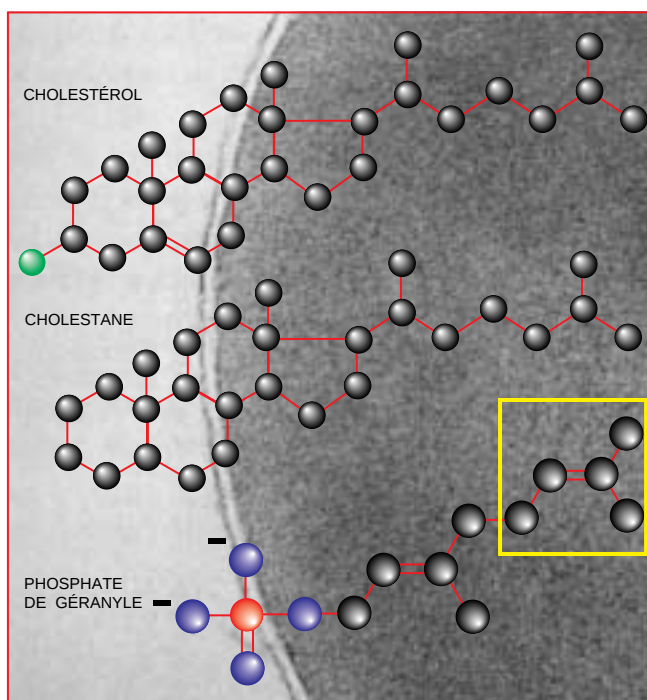
Nouveau détour de l'histoire : les pré-curseurs des géohopanoïdes furent bientôt trouvés... dans des bactéries : ces «biohopanoïdes» étudiés par Michel Rohmer constituent une classe nouvelle de lipides bactériens ; tous sont amphiphiles, avec une partie composée seulement d'atomes de carbone et d'atomes d'hydrogène, qui n'a pas d'affinité pour l'eau, et une queue hydrophile. On sait aujourd'hui, grâce à des résultats obtenus en Allemagne et à Strasbourg, que les biohopanoïdes sont localisés dans les membranes bactériennes, où ils jouent un rôle comparable à celui du cholestérol cellulaire : ils rigidifient les membranes.

DU SÉDIMENT À LA MEMBRANE

Si les biohopanoïdes sont amphiphiles, comme les phospholipides qui composent les membranes de nos cellules, ils ont la même rigidité et les mêmes dimensions que le cholestérol, qui sert de renforteur de nos membranes cellulaires. D'autre part, les géohopanoïdes ne sont pas les seuls fossiles moléculaires dont les pré-curseurs soient des terpènes : l'équipe de P. Albrecht a trouvé un grand nombre de familles moléculaires différentes, dont certaines semblent dériver de lipides membranaires bactériens connus, tandis que d'autres sont encore orphelins, comme l'étaient les géohopanoïdes au moment de leur découverte.

Connaissant la structure de ces fossiles moléculaires, les réactions susceptibles de se produire au cours de la maturation dans les sédiments et les mécanismes généraux utilisés par les êtres vivants pour synthétiser leurs constituants, on en a déduit les structures des pré-curseurs des fossiles moléculaires isolés.

Les chimistes ont alors déterminé que ces pré-curseurs, connus ou postulés, étaient tous des dérivés terpéniques, qu'ils étaient tous amphiphiles et que leur partie hydrophobe avait soit les dimensions du cholestérol, soit une longueur double, avec deux têtes polaires, hydrophiles. Ils en ont déduit – et vérifié expérimentalement – que toutes ces molécules avaient un rôle de renforcement comparable à celui du cholestérol dans les membranes modèles. Ainsi le rôle des dérivés terpéniques serait membranaire, dans tous



Trois molécules associées aux membranes (liseré sur l'image du fond) : le cholestérol (en haut), le cholestane (au milieu) et le phosphate de géranyle (en bas). Les atomes de carbone sont en noir et les atomes d'hydrogène ne sont pas représentés ; ils complètent à quatre la valence de chaque atome de carbone. En vert figure un groupe hydroxyle OH, en violet des atomes d'oxygène et en rouge un atome de phosphore. Le cadre jaune indique un motif isoprène.