

Mots croisés et anagrammes

BERNARD IQUEAUX

Un petit voyage informatique dans le monde des cruciverbistes.

Les mots croisés, pense-t-on, dérivent des mots carrés, dont l'origine se perd dans la nuit des temps : à partir de définitions qui étaient souvent données en vers, on devait trouver des mots qui se disposaient en carré et se lisaient aussi bien verticalement qu'horizontalement. Puis, au début du ^{xx}e siècle, le Britannique Arthur Wynne introduisit des cases noires, afin de simplifier la tâche des carréverbistes. Il publia sa première grille le 21 décembre 1913, dans le supplément dominical *Fun* du journal américain *New York World* : le succès fut immédiat et, comme le jeu n'était pas protégé, il se popularisa rapidement : en 1923, le Britannique Morley Adams fondait une agence spécialisée qui diffusait des grilles dans toute la presse anglo-saxonne.

En France, la première grille de «mots en croix» fut publiée dans le *Dimanche illustré*. Ces mots en croix, devenus mots croisés, furent alors repris dans divers journaux, mais ils acquirent leurs véritables lettres de noblesse en décembre 1925, lorsque les *Nouvelles littéraires* et les éditions Bernard Grasset lancèrent un concours dont le premier prix était de 10 000 francs, puis lorsque Tristan Bernard et Jean Richepin publièrent un recueil de problèmes aux subtiles définitions.

C'est alors que le mot-croisé, ou faiseur de grille, est surnommé Sphinx et que le cruciverbiste devient un Œdipe. Puis Tristan Bernard et sa collaboratrice Renée David publient des grilles assorties de magnifiques définitions, passées depuis dans le domaine public : ainsi

l'*entracte* «vide les baignoires et remplit les lavabos», le *ciel* «précède le mari». On retiendra également de Robert Lespagnol : le *pantalón* qui est une «éclipse de lune», le *sperme*, qui est une «eau de vie» ou *lo*, qui fut «victime d'une vacherie». Et l'on n'oubliera jamais Robert Scipion, qui publia dans le *Nouvel Observateur* que le *nonagénaire* fait «du vieux avec du neuf» ou que le *do* est un «demi-sommeil». Aujourd'hui Robert Scipion a créé un mouvement de Sphinx qui minimisent le nombre de cases noires (en général 10 ou 11 pour une grille de 12 cases sur 10, soit environ 9 pour cent du nombre de cases) au prix de contorsions et de chevilles, c'est-à-dire de groupe de lettres dont on s'efforce de trouver une définition (par exemple, «quartier de Béziers» pour *ezi*). Les grilles de cette tendance sont d'autant plus plaisantes qu'elles collent souvent à l'actualité.

L'autre tendance, celle de Guy Brouty, dans *Le Monde*, admet des

cases noires plus nombreuses (de 13 à 14 pour cent) mais refuse les chevilles. Le cas de la presse anglo-saxonne est particulier, car on trouve parfois dans *The Times* des grilles où les cases noires sont plus nombreuses que les cases blanches.

MOTS CROISÉS ET ANAGRAMMES

Un grand Sphinx fut Georges Pérec, qui publia notamment *La disparition*, où ne figure aucune lettre *e*. Son ami Raymond Queneau, fondateur du mouvement littéraire nommé Oulipo (*Ouvroir de littérature potentielle*), avait introduit dans son dernier roman *Le vol d'Icare*, une héroïne pour le moins originale :

«LN

Et tu t'appelles comment ?

Icare

Je ne sais plus très bien... Je ne vole plus, je nage... Et vous, Mademoiselle : Hélène ?

LN

Non, LN en deux lettres. Je suis d'origine cruciverbiste.»

Pérec, revenons-y, a produit la théorie la plus pertinente concernant la construction des grilles de mots croisés. Construire une grille est une tâche difficile si l'on veut minimiser le nombre de cases noires et si l'on veut éviter les chevilles, les mots inversés, les lettres ou parties de mots dans le désordre. Il y a 17 ans, Pérec en révélait les arcanes dans *Les mots croisés*. Précurseur éclairé, il suggérait que les mots croisés soient constitués par deux personnages distincts, celui qui construit les grilles, et celui qui laisse son imagination flotter librement, au gré

des associations évoquées par les mots, pour créer les définitions. C'est le rôle du premier personnage que nous avons imparti à l'ordinateur avec, naturellement, une modélisation et une algorithmique appropriées.

Pérec écrit également qu'une grille parfaite ne devrait comporter aucune case noire, et il cherche de telles grilles de 1, de 2, de 3, de 4, de 5 et de 6 cases de côté. Il constate que créer des grilles sans cases noires se fait presque sans effort jusqu'à quatre cases de côté, mais que la recherche d'une grille parfaite de cinq cases de côté devient très



1. Georges Pérec (en fond) fut un théoricien des mots croisés. En son honneur, pouvez-vous créer une grille de mots croisés de sept cases par cinq, sans case noire, avec les mots «Georges» en H horizontal et «Pérec» en 2 vertical (voir la solution en fin d'article)?

difficile. La constitution d'une grille parfaite de six cases de côté exige plusieurs heures de travail et «la consultation forcenée, sinon fiévreuse, d'un dictionnaire spécialisé». Il termine par une grille de sept cases de côté, avec une seule case noire au prix de «contorsions douloureuses» dont un «encrier renversé», c'est-à-dire un mot écrit de droite à gauche. Il constate également qu'il est très difficile de former des grilles de neuf cases de côté avec quatre cases noires et il lui semble impossible de placer moins de cinq cases noires pour une grille de dix cases de côté. Il admire enfin les grilles de 12 par 10 que Robert Scipion publie dans le *Nouvel Observateur* et qui ont régulièrement onze cases noires seulement.

Un autre problème voisin de la construction des mots croisés est la construction des anagrammes. Contrairement aux contrepèteries, qui sont assez faciles à créer ou à décoder pour un esprit entraîné tel celui de Luc Étienne, qui tenait la rubrique intitulée *L'album de la comtesse*, dans le *Canard enchaîné*, ou de Joël Martin, qui lui a succédé, les anagrammes constituent un exercice combinatoire des plus ardu : ce ne sont plus des sons ou phonèmes qu'il faut permuter, mais des lettres considérées comme des objets combinatoires avec lesquels on forme des mots. L'un des premiers auteurs ayant, de façon attestée, fabriqué des anagrammes est François Rabelais, qui signait Alcofribas Nasier. Plus près de nous, Boris Vian se transforma en Bison ravi. La construction d'anagrammes va nous montrer comment on peut automatiser la construction de mots croisés.

DICTIONNAIRES ET AUTOMATISATION

On a systématisé la méthode indiquée par Pérec, en utilisant un dictionnaire qui comporte, outre les féminins et les pluriels des substantifs et des adjectifs, les formes conjuguées des verbes, des noms de lieux, etc. Le dictionnaire est constitué d'une liste de mots triés dans l'ordre alphabétique. Par mot, on entend les mots de base, ou lemmes, et l'ensemble des flexions correspondantes : pluriels, féminins, formes conjuguées des verbes.

Supposons, par exemple, que nous voulions chercher dans un dictionnaire classique la signification du mot *gouleyantes*, dont la flexion est ici un féminin pluriel ; nous chercherions de fait à *gouleyant*, qui est le lemme et, dans ce cas, l'entrée du dictionnaire. La méthode que nous

avons utilisée, en collaboration avec l'ingénieur en informatique Joël Armengaud, pour constituer le dictionnaire-liste est générative : chacun des lemmes possède quatre indicateurs, qui donnent sa nature syntaxique, ainsi que les règles de formation des féminins et des pluriels. Ces indicateurs sont : la nature syntaxique (substantif, adjectif, adjectif numéral ordinal, pronom possessif, verbe, adverbe, locution, etc.) ; le genre (masculin singulier ou pluriel, féminin singulier ou pluriel, sans genre, etc.) ; la formation du féminin s'il en existe un (régulier, comme pour *grand, grande* ; ou générable, comme dans *beau, belle* ; ou encore irrégulier comme *lièvre, hase*) ; la formation du pluriel (régulier, comme *sol, sols* ; générable, comme *gavial, gavials* ; ou irrégulier comme *ail, aulx* ; invariable, etc.).

Dans le cas des lemmes qui ne sont pas des verbes, la génération des flexions se fait par programme, à partir des indicateurs décrits précédemment. Par exemple, à partir du substantif *vendeur*, on engendre *vendeurs, vendeuse, vendeuses, venderesse* (terme notarial), *venderesses*.

En ce qui concerne les verbes, nous avons mis au point un programme spécifique de conjugaison, qui tient compte du fait que, sur 9 000 verbes répertoriés, 500 seulement sont irréguliers ou défectifs. Les verbes irréguliers sont regroupés en sous-ensembles homogènes qui correspondent à une centaine de tables, tels les verbes en *ondre* (*tondre, pondre, fondre*, etc.). Les verbes réguliers du premier et du deuxième groupe ont leurs formes flexionnelles engendrées de façon totalement algorithmique, avec une décomposition en sous-groupes (*payer, je paye et je paie ; appeler, j'ap-*

pelle ; peler, je pèle ; déléguer, je délègue ; semer, je sème ; neiger, il neige ; etc.).

Quelques dictionnaires spéciaux ont été créés. Par exemple, pour les nombres de un à cent, exprimés de manière littéraire, par exemple *trente-et-un*, ou *nonante-neuf* ; pour les noms propres qui ont été arbitrairement entrés et forment une liste à part ; pour les noms de lieux (villes, régions, pays), qui forment également un dictionnaire particulier ; pour les prénoms, usuels ou exotiques.

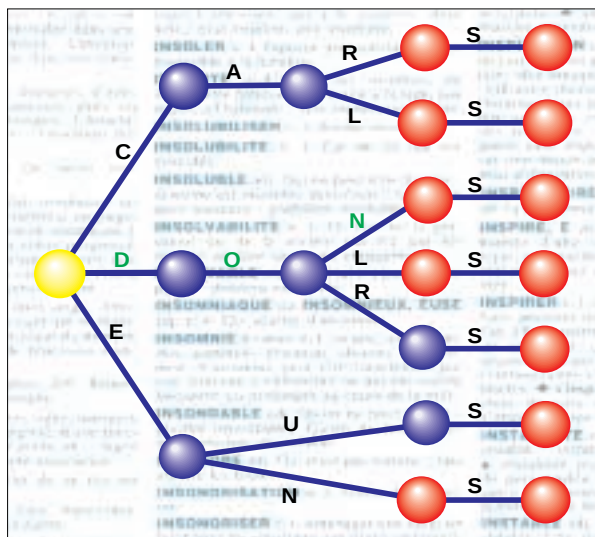
Pour obtenir le dictionnaire-liste complet, il suffit de regrouper les dictionnaires partiels précédents, de transformer les lettres minuscules en lettres majuscules et de trier la liste formée en éliminant les doublons. Naturellement on peut aussi ne retenir que certains des dictionnaires partiels, afin de constituer un dictionnaire sans noms propres ou sans formes conjuguées.

STRUCTURATION DES DONNÉES ET RECHERCHE

Tant pour la composition automatique de grilles que pour la recherche d'anagrammes, il est intéressant de savoir si un mot donné appartient au dictionnaire de référence, s'il existe des mots qui commencent par une séquence de lettres donnée, et s'il existe des mots qui se terminent par une séquence de lettres donnée.

Quand une liste de mots n'est pas triée, la seule méthode de recherche d'un mot particulier consiste à parcourir séquentiellement toute la liste. Cependant l'introduction d'un ordre de tri, alphabétique, par exemple, dans la liste des mots permet de réduire la complexité de recherche par une méthode dichotomique. Si l'on adopte une représentation arborescente de l'ensemble des mots (voir la figure 2), la résolution du problème est considérablement accélérée, parce que l'on diminue considérablement le nombre d'accès au graphe : par exemple, on n'accèdera jamais aux mots commençant par un *m* lorsque l'on cherchera un mot commençant par un *e*.

En outre, l'exploration d'un arbre nécessite moins de mémoire, car les débuts de mots ne sont présents qu'une fois. Avec Yves Gattégno, ingénieur à la Faculté d'Évry-Val-d'Essonne, et avec l'ingénieur en informatique Stéphane Moreau, nous nous sommes inspirés des travaux de Andrew Appel, de l'Université Princeton, et de Guy



2. Arbre établi pour un dictionnaire qui serait réduit à : CAR, CARS, CAL, CALS, DOL, DOLS, DON, DONS, DORS, EN, ENS, EU, EUS. Sur la figure, les nœuds terminaux sont en rouge. On a indiqué en lettres vertes le chemin qui correspond au mot DON.

Jacobson, de l'Université de Pittsburgh ; nous avons utilisé des arbres et des structures encore plus condensées, les graphes acycliques orientés, où, dans les deux cas, les chemins représentent les mots ; dans ces structures, les lettres des mots ne sont plus des nœuds, mais des arêtes du graphe, de sorte que les débuts de mots similaires ne sont représentés qu'une fois. On les parcourt à l'aide d'algorithmes de parcours récursif.

LA RECHERCHE D'ANAGRAMMES

Pour représenter le dictionnaire par un arbre, on associe une branche à chaque lettre, de sorte qu'un chemin jusqu'à un nœud terminal représente un mot du dictionnaire. Pour chercher si un mot, par exemple *abacab* est un mot du dictionnaire, on part de la racine de l'arbre et l'on se demande s'il existe des mots commençant par la lettre *a*. Si aucune arête partant de la racine n'est indexée par un *a*, il n'existe pas de mot commençant par un *a* ; si l'on trouve une arête indexée par la lettre *a*, on cherche s'il existe des chemins qui permettent de passer du *a* à un *b*. On répète alors

l'opération pour les lettres suivantes : *a*, *c*, *a*, *b*.

L'ordinateur ainsi muni d'un dictionnaire peut-il dignement succéder à Rabelais ou à Vian ? Examinons la recherche d'anagrammes du mot *esprit*. Pour résoudre ce problème, nous utilisons une méthode de recherche récursive, qui permet, quand on est dans une impasse ou que l'on a terminé d'explorer une série de solutions, de revenir en arrière jusqu'au dernier embranchement de l'arbre pour lequel la clause de vérification n'était pas encore invalidée.

Pour rechercher les anagrammes d'*esprit*, mot que l'on indique à l'ordinateur dans une variable *Mot-initial*, nous procédons de la façon suivante. Nous réservons un espace mémoire pour les résultats, le *Mot-résultat*, vide au début de l'exécution du programme. Puis nous remplissons un tableau d'une ligne et de 26 colonnes (pour les 26 lettres de l'alphabet), le *Panier*, en stockant dans chaque case le nombre d'apparitions des diverses lettres dans le mot dont on cherche les anagrammes. Par exemple, le *Panier* associé à ESPRIT est :

00001000100000010111000000.

On regarde alors quelle est la lettre correspondant au nœud exploré. Puis on

regarde dans le *Panier* la valeur de la case associée à cette lettre. Si ce nombre est nul, on ignore le sous-arbre dont le nœud exploré est la racine et on passe au nœud suivant. Si le nombre examiné dans le *Panier* n'est pas nul, on ajoute la lettre correspondant au nœud exploré à la fin de *Mot-résultat* et on diminue d'une unité, dans le panier, le nombre contenu dans la case examinée ; on applique l'algorithme général à chacun des nœuds dont le nœud exploré est le père. De plus, on vérifie si le nombre de lettres de *Mot-résultat* est égal au nombre de lettres de *Mot-initial*. Si oui, on a dans *Mot-résultat* une anagramme de *Mot-initial*, anagramme que l'on affiche (à ce stade, toutes les cases du panier sont nulles).

Dans l'exemple de la recherche d'anagrammes du mot ESPRIT, le premier sous-arbre exploré sera celui qui correspond aux mots commençant par la lettre E, avec un *Panier* où la case de E est réduite d'une unité. Une fois qu'on aura exploré ce sous-arbre tout entier, on augmentera le compteur des E dans le *Panier* (on aura alors un *Panier* identique à celui du départ) et on tentera ensuite d'appliquer l'algorithme général au sous-arbre des mots commençant par la lettre suivante dont la case du *panier* n'est pas vide (la lettre I), avec *Mot-résultat* vide et le *Panier* égal à celui de départ, à l'exception de la case de la lettre I qui aura été réduite d'une unité. Puis on répètera l'opération avec les lettres P, R, S, T. L'avantage des algorithmes récursifs est qu'ils s'appliquent de la même façon à tout un arbre ou à un quelconque sous-arbre inclus dans l'arbre total. Ils s'appliquent notamment aux feuilles terminales, et la recherche d'anagrammes, par ce biais, se fait en un seul parcours de l'arbre dictionnaire.

Les anagrammes d'ESPRIT sont ainsi obtenues en une fraction seconde sur un ordinateur à base d'un processeur *Pentium* 120 MHz de la Société *Intel* : PETRIS, PISTER, PITRES, PRIEST, PRITES, REPITS, TRIPES.

On peut aussi étendre l'algorithme de sorte que le programme trouve des anagrammes de *Mot-initial* composées de plusieurs mots. À cette fin, on ajoute simplement un test sur le nœud exploré : ce nœud correspond-il à une fin de mot valide ? Si oui, alors on ajoute un espace à la fin de mot-résultat et on applique l'algorithme général avec ce *Mot-résultat* sur l'arbre dictionnaire complet (à partir de sa racine), avec le *panier* réduit de la façon habituelle. Sinon, on ne fait rien de plus.

Notre méthode d'exploration combinatoire est exhaustive et conduit, si l'on n'y prend garde, à une inflation d'anagrammes composées de mots qui existent tous, mais dont l'association manque

CABINET A M A R R E T	CABINET AGITONS MANETTE ACERAIT ROTAMES ENTIERE TSETSES	CABINET ALUNONS MASTITE AMERRIT RATATES ENTIERE TSETSES	CABINET AZEROLE MEDITER ARASAIS ROTASSE ELEISON TESTENT	CABINET AZEROLE MEGITES ARASAIT ROYASSE ELEISON TESTENT	CABINET AZEROLE MOTIVER ANISAI RASASSE ELEISON TESTENT
CABINET AGAMIDE MARPAIN APERITE REMISAT ETAMAIT TETASSE	CABINET AGITONS MANETTE ACERAIT ROTATES ENTIERE TSETSES	CABINET ANATIDE MOLENES AMURENT RABATTE ELAITEE TASTEET	CABINET AZEROLE MEDITES ARASAIS ROTASSE ELEISON TESSENS	CABINET AZEROLE METIVES ARISAI ROSASSE ELEISON TESSENS	CABINET AZEROLE MOTIVES ANISAI RASASSE ELEISON TESSENS
CABINET AGAMIDE MARPAIN APERITE REMISAT ETAMAIT TISASSE	CABINET AGITONS MINETTE AMURENT ROTAMES ENTIERE TSETSES	CABINET ANATIDE MOLENES AMURENT RABATTE ELAITEE TESTEES	CABINET AZEROLE MEDITES ARASAIT ROTASSE ELEISON TESTENT	CABINET AZEROLE METIVES ARISAI ROSASSE ELEISON TERSENT	CABINET AZEROLE MOTIVES ANISAI RASASSE ELEISON TERSENT
CABINET AGERATE MEDISES ARASANT RAMARDE ETENDRE TESTEES	CABINET AGITONS MINETTE ARERAIT ROTATES ENTIERE TSETSES	CABINET ANATIDE MOLENES AMURENT RABATTE ELAITER TASSERA	CABINET AZEROLE MEGITES ARASAIS ROYASSE ELEISON TESSENS	CABINET AZEROLE METIVES ARISAI ROSASSE ELEISON TESTENT	CABINET AZEROLE MOTIVES ANISAI RASASSE ELEISON TESTENT
CABINET AGERATE MEDISES ARASANT RAMARDE ETENDRE TESTEES	CABINET ALIBABA MENAGER AVALERA RIRIONS ENTATES TESSERE	CABINET ANATIDE MOLENES AMURENT RABATTE ELAITER TESTERA	CABINET AZEROLE MEGITES ARASAIT ROYASSE ELEISON TERSENT	CABINET AZEROLE MOTIVER ANISAI RASASSE ELEISON TERSENT	

3. Georges Pérec n'avait pas réussi à construire une grille ayant pour potence les mots CABINET (1 horizontal) et CAMARET (1 vertical). L'ordinateur, lui, en a trouvé 28, qui sont indiquées sur cette figure.