

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 ardues : 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*, *aux* ; 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 (*paver*, *je pave* et *je paie* : *appeler*, *l'appeler*).

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.

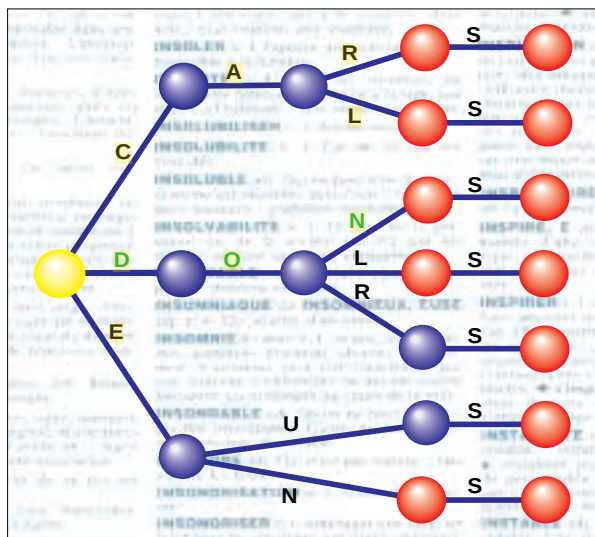
Pour obtenir le dictionnaire-liste complet, il suffit de regrouper les dictionnaires partiels précédents, de transformer les lettres minuscules en lettres majuscules et de trier la liste formée en éliminant les doublons. Naturellement on peut aussi ne retenir que certains des dictionnaires partiels, 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, DONLS, 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.