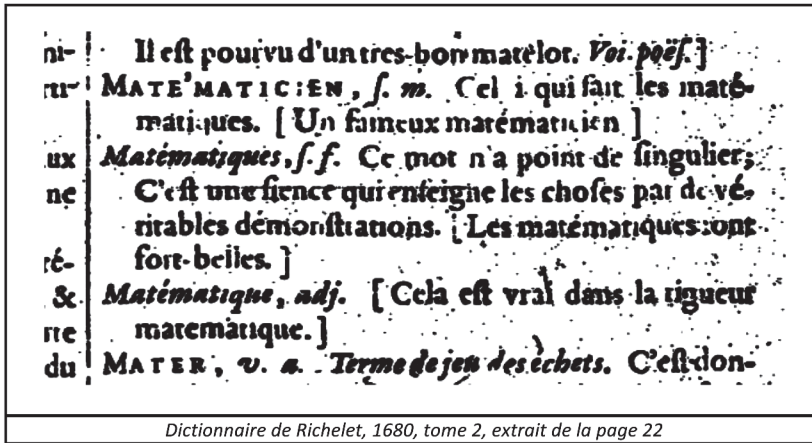


Classer des mots

Julien Moreau(*)



Dictionnaire de Richelet, 1680, tome 2, extrait de la page 22

Introduction

Nos élèves sont (ou devraient être) depuis le cours moyen familiers de deux techniques de classement : celle des entiers donnés par leur écriture décimale, celle des mots dans le dictionnaire. Faire ranger dans l'ordre croissant une liste donnée d'entiers naturels est un exercice classique à l'école primaire. Un collégien le fera d'instinct, mais sera-t-il capable de décrire le processus qu'il utilise ?

Chercher un mot dans un dictionnaire est en revanche un acquis plus incertain : nombre de collégiens, voire de lycéens, s'y montrent maladroits ou hésitants. Et regarder de près l'ordre dans lequel sont rangés les mots d'un dictionnaire réserve quelques surprises.

1. Exercices préliminaires

Phase 1 :

Ranger dans l'ordre croissant une douzaine de nombres

13569 21 573 49 109956 8741 32 9 572 42 13659 537 4876

Ranger dans l'ordre du dictionnaire une douzaine de mots

maths anglais chimie angle ange angine chameau chat bon bonne bonnet cor cri

Phase 2 :

Reprendre les listes précédentes et :

(*) julien.e.moreau@gmail.com

- ranger les nombres comme on ferait dans un dictionnaire ;
- ranger les mots comme on ferait pour des nombres.

Remarque. À ce stade, il est utile de faire observer que « est avant » joue pour les lettres (*resp.* les mots) le même rôle que « est plus petit que » pour les chiffres (*resp.* les nombres).

Phase 3

Quand la phase 2 est réussie, on peut demander aux élèves d'explicitier le processus (pour ne pas dire l'algorithme) utilisé dans les deux cas. Ce n'est pas si facile, car appliquer d'instinct une règle ou même savoir la transposer à une nouvelle situation ne rend pas pour autant capable de décrire la démarche suivie. Essayez pour voir de détailler l'accord des participes passés ou les opérations à faire pour effectuer un créneau !

Il n'est pas question de faire mettre en forme un algorithme, mais d'obtenir une description cohérente dans le langage de tous les jours, quelque chose comme ce qui suit :

Pour les nombres

- * mettre d'abord les nombres à un seul chiffre, s'il y en a, puis ceux à deux chiffres, s'il y en a, puis...
- * à nombre de chiffres égal, mettre d'abord les nombres dont le premier chiffre en partant de la gauche est le plus petit, puis ...
- * à nombre de chiffres égal et à premier chiffre égal, mettre d'abord les nombres dont le second chiffre est le plus petit, puis ...

Pour les mots

Par commodité, nous écrivons le signe « < » pour « est avant ». Ainsi

$$a < b < c < \dots < z.$$

- * Pour comparer les mots de même longueur, on utilise la même convention que pour les entiers de même longueur. Ainsi : *abri < amer < amie < azur < baba < banc.*
- * Comparer deux mots de longueur différente est plus laborieux, car on n'utilise pas la même règle que pour les entiers, qui est « le plus court d'abord » ; sinon *or* serait avant *argent*. On coupe le plus long à la taille du plus court en supprimant sa ou ses dernières lettres ; l'ordre des deux mots est alors celui du mot court et du mot raccourci (qui bien sûr n'a pas forcément de sens), sauf si le mot court est le début du mot long, auquel cas le mot le plus court est mis devant. Ainsi *amateur < amer* (parce que *amat < amer*) et *amer < amertume*.

Question : À quoi pourrait servir un dictionnaire dans lequel les règles de classement seraient celles des entiers ?

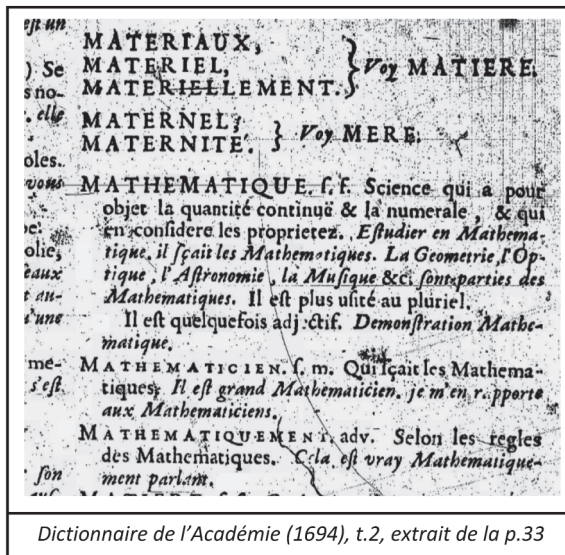
Réponse : À faire des mots croisés, et de tels dictionnaires existent.

2. Le classement des mots dans les dictionnaires

Une fois énoncés les principes généraux que nous venons de voir, les élèves estimeront en général en avoir terminé. Il restera à leur faire sentir qu'on est encore loin du compte.

Parenthèse historique

Les dictionnaires n'ont pas tous, au début, rangé les mots dans l'ordre alphabétique. Le *Dictionnaire mathématique* d'Ozanam (1691) les regroupait par rubriques (par exemple : « Arithmétique vulgaire », « Géométrie spéculative », « Navigation »). Le premier dictionnaire de l'Académie française (1694) les classait par familles, elles-mêmes classées dans l'ordre alphabétique (voir l'image ci-après).



La genèse de ce premier dictionnaire de l'Académie française mérite d'être contée. En projet depuis la création de cet auguste corps (1635), le dictionnaire n'était encore en 1674 qu'une vague ébauche. Rechignant à se presser, mais ne voulant pas que quiconque la précède, l'Académie obtint du roi en 1674 l'exclusivité pour vingt ans de la publication en France.

Hélas, ce fut Richelet qui en 1680 publia à Genève le premier dictionnaire du français (qui, lui, suivait strictement l'ordre alphabétique). Furieuse, l'Académie fit saisir une bonne partie des exemplaires du livre de son rival heureux.

L'histoire ne s'arrête pas là : l'académicien Furetière, excédé de l'inertie de ses collègues, entreprit de rédiger à lui tout seul un dictionnaire et publia en 1684 un avant-projet qui lui valut d'être exclu de l'Académie. Il eut une éclatante revanche posthume : son dictionnaire, paru à Rotterdam en 1690, deux ans après sa mort, devint pour plus d'un siècle l'ouvrage de référence, tandis celui de ses ex-collègues, publié en 1694, sombrait vite dans l'oubli.

Le problème des signes diacritiques

- La principale difficulté dans le classement des mots est liée à la présence de signes « diacritiques⁽¹⁾ », c'est-à-dire de signes modifiant la prononciation d'une lettre : cédille, accents, tréma.
- Ces signes n'existent pas en anglais, mais le français est loin d'être la seule langue à en posséder : ainsi les Allemands ont l'*Umlaut*, un tréma qui, mis sur *a*, *o* ou *u*, en change le son, les Espagnols ont la tilde ~ qui se met sur le *n*, les Portugais ont un ç.
- Remarque : Les ligatures æ et œ ne posent aucun problème : les mots les contenant sont classés comme si les deux lettres étaient séparées.

La règle de base

- Comment placer, par exemple, *dû* et *dûment* par rapport à *du*, *dual*, *duc*, *dune*, *dupe*, *dur* ? N'importe quel dictionnaire nous donne l'ordre :

du < dû < dual < duc < dûment < dune < dupe < dur.

Cela nous permet déjà d'éliminer une hypothèse : le *û* n'est pas considéré comme une lettre supplémentaire à insérer dans l'alphabet. S'il l'était, il serait après *u*, puisque *du < dû*, et on aurait alors *dune < dûment*. Tout s'est passé, en fait, comme si dans le classement on avait oublié l'existence de l'accent circonflexe. Nous pouvons donc conjecturer une règle :

Règle n° 1 : *Le classement des mots se fait indépendamment des signes diacritiques.*

- On la vérifie aisément sur quelques exemples :

glaçage < glace < glaçon ; coup < coupé < couper ; rapace < râper < rapine ; etc.

Remarque. Les dictionnaires espagnols traitent *ñ* comme une lettre distincte de *n*, située entre *n* et *o*. On pourrait penser qu'il aurait été plus rationnel de les imiter et de considérer les lettres affectées d'un signe comme des lettres à part entière. L'alphabet s'écrirait : *a, à, â, b, c, ç, d, e, é, è, ê, ë, f, g, h, i, î, ï, j, ...*

Mais ce serait introduire le divorce dans beaucoup de familles de mots. L'usage français est en effet incertain en matière d'accentuation : on aurait été amené à séparer, par exemple, *règle* et *règlement* de *régler* et *réglementaire*.

Les trous de la règle n° 1

- La règle que nous venons d'énoncer marche dans presque tous les cas. Hélas, dans un dictionnaire comportant quelque soixante mille entrées, ce « presque tous » laisse échapper encore pas mal de choses.

Cette règle est en défaut si et seulement si on doit classer deux mots qui ne diffèrent que par les signes diacritiques.

- Prenons quelques exemples. Un quelconque dictionnaire nous donne : *cote < côte*, *coupe < coupé*, *du < dû*, *la < là*, *mat < mâ*, *ou < où*, *pale < pâle*, *tache < tâche*.

Cela nous permet d'énoncer :

(*) Du grec διακριτικός, « qui sert à séparer ».

Règle n° 2 : *Si deux mots ne diffèrent que par des signes diacritiques et si l'un n'en a pas, il est placé avant l'autre.*

- Voilà qui élimine beaucoup de cas litigieux. Mais que dire lorsque les deux mots ont l'un et l'autre des signes diacritiques ? Voyons encore quelques exemples :

chassé < châsse ; maçon < mâcon ; pâte < pâté

Ils nous permettent de poser une troisième règle :

Règle n° 3 : *Étant donné deux mots ne différant que par des signes diacritiques, on regarde la première lettre (à partir de la gauche) pour laquelle ils diffèrent ; si dans un des mots cette lettre n'a pas de signe diacritique, il est placé en premier.*

- Ces trois règles permettent de traiter la plupart des cas, mais pas tous : comment justifier, par exemple *gène < gêne* ou *péché < pêche* ? Le second exemple est le plus surprenant, car le mot à deux signes diacritiques est placé avant celui qui n'en a qu'un. On utilise en fait la règle que voici.

Règle n° 4 : *Étant donné deux mots ne différant que par des signes diacritiques, on ne tient compte que de la première lettre (à partir de la gauche) pour laquelle se présente une différence ; le mot placé en premier est celui où cette lettre est la plus « petite » au sens de la convention suivante :*

- *pas d'accent < accent aigu < accent grave < accent circonflexe < tréma ;*
- *pas de cédille < cédille.*

Notons que les règles n° 2 et n° 3 sont des cas particuliers de la règle n° 4.

Alors, on a fini ?

- **Mais non, voyons ! On a oublié les mots composés.**

Les mots composés

- Les mots composés s'écrivent le plus souvent en accolant les composantes ou en les séparant par un trait d'union. L'usage en la matière est passablement arbitraire : pourquoi écrit-on *portemanteau* et *porte-jupe*, *baisemain* et *baise-en-ville* ? Il est en outre fluctuant : on trouve *contrepied* et *contre-pied*, *millefeuille* et *mille-feuille*, etc.

- Quand les composantes sont accolées, le mot est tout bonnement placé selon les règles habituelles. Mais la présence d'un trait d'union (ou de plusieurs, comme dans *cul-de-basse-fosse*) pose un problème. Voyons comment un dictionnaire s'en tire :

porte < porte-fenêtre < portefeuille < porte-parole < porter < porte-savon < porteur.

La situation est limpide : on a fait comme si le trait d'union n'existait pas. Il est facile (et souhaitable) de vérifier sur d'autres exemples, ce que je laisse aux soins du lecteur.

- Il reste une dernière catégorie de mots composés, ceux où les composantes ne sont ni accolées, ni jointes par un trait d'union, mais séparées par un intervalle. Il y a parmi eux beaucoup de termes d'origine étrangère : *a priori* ; *ave maria* ; *casus belli* ; *de visu* ; *in petto* ; *nota bene* ; *persona grata* ; *statu quo*, etc.

Mais il y en a aussi qui sont bien de chez nous, comme : *ayant droit, compte rendu, pomme de terre* et ... *trait d'union*.

Voyons le principe de classement : *âpre* < *a priori* < *apte* ; *casuiste* < *casus belli* < *catalan* ; *compter* < *compte rendu* < *compteur*.

On subodore sur ces trois exemples que l'intervalle entre deux composantes n'est pas plus pris en compte que le trait d'union.

Règle n° 5 : *Les mots composés sont traités comme si leurs composantes étaient accolées.*

Cette fois, c'est bien fini !

– Erreur ! On a oublié les homographes.

Deux mots sont homographes si leur écriture est exactement la même, par exemple *cent* monnaie et *cent* nombre, *corner* verbe et *corner* au football, *cousin* parent et *cousin* moustique, *grue* appareil et *grue* oiseau, le *mousse* et la *mousse*, le *pic* oiseau, le *pic* outil et le *pic* montagne, le *tour* et la *tour*, la *vanne* robinet et la ou le *vanne* plaisanterie (est souligné le mot mis en premier dans le Robert).

On voit que lorsqu'il y a une nette différence d'usage, le plus courant est mis en premier. Sinon je crains fort que le choix ne soit fait au hasard. Les lexicographes, gens sérieux entre tous, joueraient-ils à pile ou face ?

3. Le classement des mots par les logiciels de tri

Il est commode, pour classer dans l'ordre alphabétique une liste de mots, de les disposer sur une colonne d'un tableau de *Word* ou d'*Excel*, par exemple, et d'appliquer tout bêtement la commande « Trier ».

Prenons par exemple ce vers⁽²⁾ de Racine :

Le jour n'est pas plus pur que le fond de mon cœur.

Le tri donne impeccablement : *cœur de est fond jour Le le mon n' pas plus pur que*. Faut-il pour autant conclure que tout est pour le mieux dans le meilleur des mondes ? Ce serait un peu rapide : l'exemple ci-dessus ne comporte ni signes diacritiques ni mots composés. Ce serait aussi oublier que dans tout Éden il y a un serpent. Nous allons même ici en débusquer deux.

Les mots à signes diacritiques

• Voyons ce qu'il en est quand on prend des mots à signes diacritiques, en excluant provisoirement les mots composés.

La liste ci-dessous est rangée dans l'ordre du dictionnaire :

cote côte coupe coupé gène gêne maçon mâcon pale pâle péché pêche tache tâche

La voici dans l'ordre que donne la commande « Trier » de *Word* :

cote côte coupe coupé gène gêne mâcon maçon pale pâle pêche péché tache tâche

(2) *Phèdre*, acte IV, scène 1. C'est un exemple rarissime de vers comptant treize mots en douze syllabes.

La liste n'est pas exactement la même. L'ordre, abstraction faite des signes diacritiques, est respecté, la règle « *si deux mots ne diffèrent que par des signes diacritiques et si l'un n'en a pas, il est placé avant l'autre* » l'est aussi, mais il y a deux interversions.

• Comprendre ce qui s'est passé n'est pas immédiat. La recherche est simplifiée par deux remarques :

- 1) *Word* se fichant royalement du sens des mots, on peut travailler sur des mots qui ne sont pas dans le dictionnaire ;
- 2) on peut donc supprimer les parties qui sont identiques (et placées au même endroit) dans les deux mots à comparer.

Prenons le problème posé par *péché* et *pêche*. Il nous amène à comparer *ée* et *êe* ; la comparaison directe de ces deux « mots » par tri donne *êe* < *ée*, ce qui recoupe bien le classement fait sur les deux mots complets, mais ne l'explique pas. Pour l'expliquer, plaçons-nous dans un cadre plus général, en considérant la liste de toutes les variantes de *ee*, rangées conformément aux règles du dictionnaire dans la ligne supérieure, rangées par le tri de *Word* dans la ligne inférieure.

↓ liste rangée dans l'ordre du dictionnaire ↓			
<i>ee ée èe êe ëe</i>	<i>ée éé èè êê ëë</i>	<i>èe èé èè èê èë</i>	<i>êe êé êè êê êë</i>
<i>ee ée èe êe ëe</i>	<i>éé éé èè éé éé</i>	<i>eè eè èè êè èè</i>	<i>eê eê èè êê èê</i>
↑ liste rangée par la fonction « tri » de <i>Word</i> ↑			

On voit que l'on passe d'une ligne à l'autre en permutant dans chaque « mot » les deux lettres qui le composent : la lettre de gauche passe à droite et la lettre de droite passe à gauche.

Le classement *Word* est donc fait selon la règle n° 4 énoncée précédemment, avec cette différence que *le mot « gauche » est remplacé par le mot « droite »*.

On peut ainsi proposer une règle :

Étant donné deux mots ne différant que par des signes diacritiques, la fonction « Trier » ne tient compte pour les classer que de la première lettre (à partir de la droite) pour laquelle se présente une différence.

Le classement par *Word* du couple *maçon mâçon* confirme la validité de cette règle. Le lecteur sceptique pourra vérifier que *Word*, à la différence des dictionnaires, place en conséquence *châsse* avant *chassé*.

Les mots composés

Eux aussi nous réservent une surprise. Prenons dans un dictionnaire la liste suivante : *abatis abat-jour abattoir abattre âpreté a priori avaler à-valoir avanie*.

Triée par l'ordinateur, elle devient :

a priori abatis abat-jour abattoir abattre âpreté avaler à-valoir avanie

Le logiciel respecte la place des mots à trait d'union, mais il case « *a priori* » en tête. On peut penser qu'il considère l'intervalle entre le *a* et le *p* comme une lettre située

en tête de l'alphabet, en tout cas avant *b*. Pour le vérifier, il suffit d'introduire dans la liste le mot *Aa*, nom d'un fleuve côtier bien connu des cruciverbistes : le tri le met en seconde position derrière *a priori*. Cela nous permet de conclure :

Pour le logiciel, l'intervalle est considéré comme la lettre initiale de l'alphabet.

Moralité. Utiliser la fonction « tri » d'un ordinateur donne une méthode rapide et agréable de classer dans l'ordre alphabétique une liste de mots. Mais il faut corriger ce tri pour les couples de mots ne différant que par les signes diacritiques et les mots composés sans trait d'union.

Remarque. J'ai dans cet article utilisé le tri de *Word 2007*. Mais celui d'*Open Office* donne les mêmes résultats.

Conclusion

Il n'y a pas beaucoup de mathématiques dans tout cela, direz-vous. En un sens, d'ailleurs, c'est exact : il serait vain de chercher à rattacher cet article à une quelconque tranche du programme des classes. Il n'empêche que son contenu n'est autre qu'une étude informelle d'algorithmes de tri.

En outre, la méthode utilisée tout au long de ce texte a été la suivante : rassembler un lot d'exemples, faire à partir de là une conjecture, la valider sur d'autres exemples. Cette façon d'agir ressemble fort au type de travail effectué en sciences expérimentales... et à celui que l'on fait en mathématiques avant de passer au stade d'une démonstration mise en forme.

Alors pourquoi ne pas essayer avec vos élèves ce thème, qui a de surcroît le mérite d'être interdisciplinaire, donc dans la droite ligne du discours officiel ?

Post-scriptum

Au terme de ce travail, un doute me prend. Et si j'étais resté prisonnier de la tradition scolaire ? Et si j'avais obstinément tourné le dos à la modernité ? Avec des élèves plus habiles à pianoter devant leur console qu'à tourner les pages d'un livre, ne conviendrait-il pas d'inverser l'ordre des facteurs et de commencer par analyser le tri selon *Word* ou *Open Office* avant d'aborder le dictionnaire, cet outil du passé ? À vous de juger !