

La partie «mobile» de la phrase est formée des lettres des nombres associés aux lettres pouvant figurer dans des nombres. Le corps d'une phrase étant choisi, c'est cette partie mobile que nous recherchons pour écrire une phrase réflexive. Pour la phrase du titre et du chapeau, la partie mobile est : *huitsix-cinqvingttroisdeuxtroisquatredixsept-seizehuit-cinqseptdixonzevingtquatrequatorzetroisseptquatre*.

Le problème a quelques variantes. Ainsi, en français, on peut considérer les caractères e, é, ê, ë et ù comme une seule lettre ou comme cinq lettres différentes. Le choix d'un classement ou d'un autre n'est toutefois pas très important : en français, aucun nombre ne contient de lettre accentuée, à l'exception de zéro, qui ne peut apparaître (si l'on indique que l'on a zéro fois une certaine lettre, cette lettre apparaît une fois, et le zéro disparaît!). Ainsi, la distinction des lettres accentuées alourdit seulement la partie fixe de la phrase avec des nombres qui ne changeront jamais.

La phrase peut aussi mentionner le nombre d'espaces, de signes de ponctuation, de traits d'union, voire le nombre total de ses lettres. Avec les signes de ponctuation, on augmente seulement la partie fixe de la phrase : leur quantité n'est pas modifiée quand on change les nombres. En revanche, compter les traits d'union ajoute des lettres à la partie mobile, car tous les nombres n'ont pas de trait d'union. De même si l'on compte les espaces : trente et un a deux espaces alors que d'autres nombres n'en ont aucun. Enfin, le nombre total de lettres change évidemment quand on modifie un nombre : pour les nombres inférieurs à 100, il va de 2 (un) à 19 (quatre-vingt-quatorze). Dans chacune de ces trois variantes, on augmente d'une unité la quantité de nombres à trouver pour définir une phrase, ce qui complique un peu le problème.

On peut aussi rechercher les phrases réflexives les plus concises possible : plutôt que d'indiquer qu'il y a, par exemple, «un k» ou «un w», on ne fait pas figurer ces lettres, qui n'apparaissent qu'à l'endroit de la phrase où l'on en mentionne le nombre. Il est difficile d'être plus concis que la phrase suivante : «Trois a, deux c, six d, dix e, deux h, huit i, deux n, trois o, quatre p, quatre q, cinq r, sept s, dix t, sept u et sept x.»

Enfin, en hommage à Georges Perec, voici une phrase réflexive qui ne contient pas de e. Elle est un peu alambiquée, car elle ne peut contenir que les nombres un, trois, cinq, six, huit, dix, dix-huit, vingt, vingt-trois, vingt-cinq, vingt-six et vingt-huit. Le nombre sans e suivant est un million ! «Ha, l'on trouva la solution ; il

Les éléments d'une phrase réflexive

On choisit d'abord le **corps** de la phrase, c'est à dire les lettres qui ne sont pas dans les nombres associés aux lettres :

cetitrecontientabcdefghijklmnpqrstuvwxyet z

Puis on détermine la **partie fixe** de la phrase, le corps plus les lettres des nombres associés aux lettres qui n'apparaissent dans aucun nombre (b, j, k, l, m, w et y) :

cetitrecontientaunbcdefghiunjunkunlmnopqrstuvunwxunyet z

Il ne reste qu'à déterminer la **partie mobile** de la phrase, les nombres associés aux lettres qui apparaissent dans les nombres :

Ce titre contient quatre a, un b, cinq c, cinq d, dix-neuf e, deux f, un g, deux h, treize i, un j, un k, un l, un m, seize n, trois o, quatre p, sept q, sept r, sept s, quinze t, dix-huit u, un v, un w, six x, un y et quatre z.

rit, titra tout, sans tri, sur un trait gras : cinq c, huit a, trois g, six d, vingt-cinq i, six h, dix n, cinq l, cinq q, huit o, dix s, dix r, dix u, dix-huit t, huit x, trois v».

DES CHIFFRES ET DES LETTRES

Comment écrire de telles phrases ? Pour un corps donné, on pourrait écrire toutes les phrases possibles en utilisant systématiquement tous les ensembles possibles de 26 nombres ; on vérifierait ensuite seulement si chacune de ces phrases décrit correctement son contenu. Cette méthode est toutefois inutilisable : le nombre de situations possibles est infini, et même en restreignant le choix aux nombres inférieurs à 30, par exemple, le plus puissant des ordinateurs n'en viendrait pas à bout en un temps raisonnable.

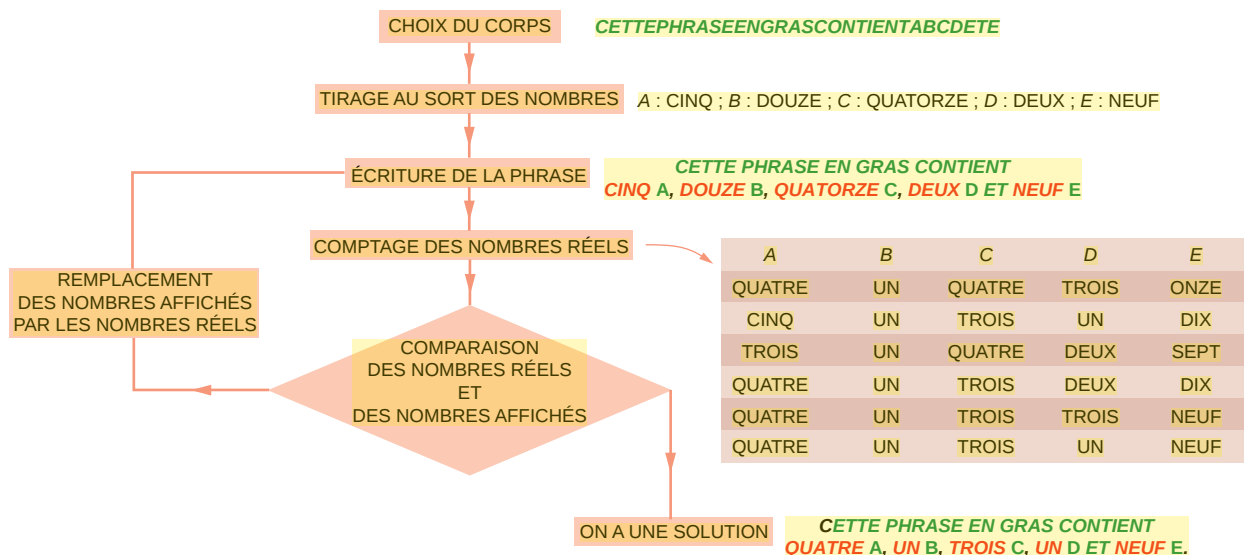
En français, on peut réduire le nombre de combinaisons : nous avons vu que seulement 19 lettres peuvent figurer dans un nombre. En outre, les lettres g et v apparaissent toujours ensemble, et seulement dans le nombre vingt, le nombre de g et de v qui apparaissent dans des nombres sont toujours les mêmes. Les nombres associés au g et au v figurent toutefois dans la partie mobile : ils peuvent tous les deux changer, bien que ces changements soient liés. Ainsi, un corps étant donné, nous n'avons que 18 nombres à déterminer. C'est encore trop.

Heureusement, Douglas Hofstadter, dans son livre *Ma Thémagie*, a décrit une méthode plus efficace pour créer des phrases réflexives. Pour chaque lettre de l'alphabet, on nomme «nombre réel» le nombre d'apparitions de cette lettre dans la phrase et «nombre affiché» celui qui, dans la phrase, est associé à cette lettre. Ainsi, dans «cette phrase contient quatre a», le nombre affiché pour a est quatre et le nombre réel est trois.

Le corps de la phrase étant choisi, on commence alors en tirant au sort les 26 nombres affichés que l'on cherche. Si, pour toutes les lettres, le nombre affiché est égal au nombre réel, on tient une solution. Si ce n'est pas le cas, on remplace pour chaque lettre son nombre affiché par son nombre réel, qui devient ainsi le nouveau nombre affiché. On poursuit le processus en vérifiant à chaque itération si les nombres affichés sont ou non égaux aux nombres réels.

Deux cas peuvent se produire : soit on arrive à une solution, soit on retrouve une situation déjà rencontrée et l'on tombe dans une boucle. On pourrait détecter les boucles en gardant les situations déjà rencontrées, mais cela obligerait à conserver en mémoire beaucoup trop de situations et nécessiterait un trop grand nombre de comparaisons. Il est plus simple de limiter le nombre d'itérations permises : si l'on dépasse cette limite, on arrête, on refait un tirage pour les nombres affichés et l'on recommence les itérations. On fixe aussi une limite au nombre de tirages de départ. Si on la dépasse, on arrête les essais pour ce corps. Ces nombres maximaux d'itérations et de points de départ ont une grande influence sur le succès de la méthode et sur le temps de calcul.

Pour un corps donné, il n'existe pas toujours une phrase réflexive, même dans le cas simple où l'on s'intéresse aux apparitions d'une seule lettre. Ainsi la phrase : «Cette courte phrase contient en tout six e» est incorrecte puisqu'elle contient sept e. Changeons alors le six en sept : «Cette courte phrase contient en tout sept e.» Cela ne va toujours pas puisqu'il y en a maintenant huit. Changeons le sept en huit : «Cette courte phrase contient en tout huit e.» Et nous revenons à sept e. Si l'on met neuf, on aura seulement huit e : il n'y a pas de solution.



Un algorithme permet de trouver des phrases réflexives dont on a initialement choisi le corps : on tire au sort les nombres associés aux lettres, on écrit la phrase, et on compare ces nombres affichés avec les nombres réels. Si des nombres réels sont différents des nombres affichés, il les remplacent, et on recommence jusqu'à ce que l'on ait trouvé une solution. Dans l'exemple, volontairement limité aux cinq premières lettres de l'alphabet, on obtient une solution après six itérations seulement (les

valeurs des nombres réels à chaque itération sont donnés dans le tableau à droite). Avec tout l'alphabet, plusieurs milliers d'itérations sont généralement nécessaires, et l'on ne trouve pas nécessairement de solution, même s'il en existe une. Une amélioration de la méthode consiste à ne remplacer, à chaque étape, qu'un seul des nombres affichés par le nombre réel de la lettre correspondante : on bouleverse ainsi moins la phrase à chaque étape, et on trouve plus fréquemment une solution.

Par cette méthode, nous trouvons quelquefois des solutions. Son succès est surprenant, car la phrase est bouleversée à chaque itération : la valeur de la plupart des nombres appartenant à la partie mobile est modifiée à chaque étape. Chaque phrase est donc très différente de la précédente, et l'on ne s'approche pas régulièrement de la solution (si elle existe) : on peut successivement s'en approcher et s'en éloigner.

Toutefois, ce procédé a un avantage important : si l'on arrive à une situation où l'ensemble des lettres des nombres réels est identique à l'ensemble des lettres des nombres affichés, on obtient la solution à l'étape suivante. Supposons que nous ayons dans la phrase du titre et du chapeau « quatre t », au lieu de vingt-quatre, et « vingt-sept x » au lieu de sept, les autres nombres étant les mêmes. Les nombres réels donneront, à l'étape suivante, « vingt-quatre t » et « sept x », ce qui est la solution.

UN SEUL REMPLACEMENT

J'ai amélioré cette méthode. Au lieu de remplacer à chaque itération tous les nombres affichés par les nombres réels correspondants, le programme choisit une seule lettre dont le nombre affiché est différent du nombre réel, et il ne remplace que le nombre affiché de cette lettre par son nombre réel. Cette nouvelle méthode doit préserver une qualité essentielle de la précédente : quand l'ensemble des lettres des nombres réels est le même

que l'ensemble des lettres des nombres affichés, on doit obtenir la solution à l'étape suivante. Sinon, on risque de passer à côté d'une solution évidente. J'ai donc ajouté une vérification : le programme détermine l'ensemble des lettres des nombres réels et celui des lettres des nombres affichés qui en diffèrent. Si ces deux ensembles sont identiques, l'ensemble des nombres réels est une solution. Pour ne pas trop allonger le temps de calcul, le programme ne fait cette comparaison que si l'écart entre les nombres réels et les nombres affichés (l'écart est la somme des valeurs absolues des différences entre les nombres réels et les nombres affichés) n'est pas trop grand.

Cette amélioration est performante : elle permet de trouver plus de solutions. Cela est dû à ce que la phrase est moins bouleversée à chaque itération, puisque l'on ne change qu'un seul nombre affiché : la partie mobile est plus stable.

Pour trouver de nombreuses phrases, j'ai introduit dans le programme un générateur de corps de phrases. J'ai défini d'une part un « noyau », suite de mots commune à tous les corps, et j'ai défini d'autre part plusieurs endroits, dans le noyau, où le programme peut insérer un « ajout », groupe de mots pris parmi plusieurs possibilités liées à chaque endroit. La présence ou l'absence de ces ajouts ne change pas le sens du corps : par exemple, « cher lecteur », « chers lecteurs », « ami lecteur » en tête de phrase ou « pour terminer », « finalement », « pour finir » avant le z final. En combinant une dizaine de

possibilités d'ajout, on crée de nombreux corps. À partir d'un même noyau, le programme engendrait 12 288 corps différents, ce qui pouvait donner naissance à environ 9 000 phrases réflexives.

PLURALITÉ DE SOLUTIONS

Le même corps donne parfois plusieurs solutions. Pour le cas de deux caractères, nous avons vu la phrase : « Cette phrase contient exactement cinq n et trois o. » Il existe d'autres solutions sur le même corps, par exemple : « Cette phrase contient exactement quatre n et trois o. », « Cette phrase contient exactement cinq n et deux o. », « Cette phrase contient exactement quatre n et deux o. »

Ce phénomène est fréquent, même quand on indique les nombres d'apparitions des 26 lettres de l'alphabet : pour les corps ayant au moins une solution, le nombre moyen de solutions est d'environ 1,5. Le record actuel est de dix phrases différentes construites sur le même corps. Voici l'une de ces dix phrases : « Vous avez bien raison de compter toutes les lettres de cette phrase avant de la croire quand elle vous affirme qu'elle contient quinze a, deux b, sept c, sept d, quarante-trois e, quatre f, cinq g, trois h, vingt et un i, un j, un k, huit l, quatre m, vingt-cinq n, treize o, sept p, douze q, vingt et un r, quinze s, trente-quatre t, vingt-quatre u, neuf v, un w, trois x, un y et pour terminer six z. »

À partir d'un point de départ, on peut soit arriver à une solution, soit tourner indéfiniment dans une boucle.