

REVUE DE STATISTIQUE APPLIQUÉE

A. VESSEREAU

Statistique et français

Revue de statistique appliquée, tome 25, n° 1 (1977), p. 5-11

<http://www.numdam.org/item?id=RSA_1977__25_1_5_0>

© Société française de statistique, 1977, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/legal.php>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

STATISTIQUE ET FRANGLAIS

A. VESSEREAU

Ce que l'on conçoit bien s'énonce clairement
Et les mots pour le dire arrivent aisément.

N. BOILEAU —

“*Ce que l'on conçoit bien. . .*”. Dans le domaine de la statistique, il arrive que cela ne soit pas tellement facile.

“*Et les mots pour le dire. . .*”. Ce qui était aisé à l'époque de Boileau n'est pas toujours à l'époque des “sciences et des techniques”.

Les remarques qui suivent sont dédiées à l'auteur de l'Art Poétique.

*

* * *

Je serais navré que nos amis anglais ou anglophones prennent le moindre ombrage de cette note. Il n'est nullement dans les intentions de l'auteur de revendiquer la suprématie de sa langue sur la leur ⁽¹⁾. L'anglais et le français ont leur génie propre, et cela n'a pas empêché de nombreux échanges au cours des âges. Un seul exemple suffira : le vieux mot français “reille”, dérivé du latin “regula” — qui signifiait “barre”, ou “poutrelle”, est depuis longtemps abandonné. Il a été anglicisé pour donner “rail”, puis réincorporé à la langue française, avec l'orthographe anglaise mais la prononciation française — en souvenir inconscient peut-être de l'ancien “reille”.

Il serait absurde de fixer quelque langue que ce soit dans son état actuel. L'évolution est permanente et nécessaire ; toute langue — sauf si elle est “morte” — rejette des “déchets” ; pour désigner des concepts nouveaux, elle élargit la signification de termes existants, ou en invente de nouveaux.

Il n'est donc pas condamnable en soi que la langue française adopte des termes anglais, ou dérivés de l'anglais. L'important est que l'évolution ne s'effectue pas de façon anarchique, et n'aboutisse pas à la longue à la contamination

(1) On aurait aussi bien pu mettre, en exergue à cette note : “Words, words, words, . . .” (W. Shakespeare — Hamlet).

d'une langue par une autre. Assez curieusement, le retour au "bon sens", après un faux-pas souvent imputable à la mode, s'effectue parfois de lui-même. Nos grands-parents, lorsqu'ils décidaient d'aller prendre quelque repos sur la Côte d'Azur, et s'ils en avaient les moyens, voyageaient en "sleeping", (ou en "sleeping-car") ; ils utilisent maintenant un wagon (ou une voiture)-lit. Mais n'est-il pas préférable d'éviter de tels . . . allers et retours.

*
* *

Beaucoup de concepts statistiques (et cela est vrai pour bien d'autres sciences ou techniques) sont d'origine anglaise. Pour les désigner, les anglais ont, soit utilisé un mot existant en lui donnant une acception nouvelle, soit créé un mot nouveau à consonance anglaise bien évidemment.

Dans le premier cas, à côté d'un (ou plusieurs) sens généraux, le même mot (ne pas confondre avec des mots identiques, mais de racines différentes) se trouve avoir un, ou plusieurs, sens spécifiques. Cela ne présente pas de graves inconvénients, le contexte permettant généralement d'écarter les fausses interprétations. Sans qu'il soit besoin de rappeler le turc du "Bourgeois Gentilhomme" (quelle "langue admirable que ce turc" . . . qui dit "tant de choses en deux mots") on peut affirmer qu'en aucune langue il n'y a correspondance biunivoque entre mot (linguistique) et signification (sémantique) ; en termes plus modernes on dirait qu'il n'y a pas bijection entre signifiant et signifié. Cela entraîne des difficultés dans l'"application" du français à l'anglais, mais celles-ci sont rarement insurmontables.

Dans le deuxième cas (création d'un mot anglais) faut-il se priver d'inventer un mot nouveau à consonance française ?

Une remarque doit être faite ici sur certains pièges qui se présentent lors de la recherche d'un équivalent français à un mot anglais. Il arrive que le "mot anglais" existe déjà en français, avec la même orthographe (à défaut de la même prononciation) mais avec un sens très différent. "Accuracy" a pour équivalent français précision, mais "precision" (anglais) doit se traduire par fidélité. Au couple de concepts, fondamentaux en matière de mesures, "accuracy-precision" correspond en français le couple "précision-fidélité". C'est un exemple banal, mais il en existe bien d'autres.

Ces quelques considérations d'ordre général nous amènent à penser que, dans la recherche de l'équivalent français d'un mot anglais, trois situations peuvent être envisagées.

1) Existe-t-il un mot français équivalent, ou suffisamment équivalent, pour qu'à l'usage il n'y ait pas ambiguïté ?

2) Sinon, peut-on créer un mot à consonance française, qui pourra, si nécessaire, être calqué sur le mot anglais ?

3) L'adoption telle quelle du mot anglais (avec quelle prononciation ?) ne devrait être retenue que si l'on n'a pas pu répondre positivement, et de manière satisfaisante, aux questions précédentes. Il s'agirait d'une adoption provisoire, le mot anglais étant placé "entre guillemets" ; l'adoption définitive pourrait intervenir plus tard.

Quelques exemples historiques illustreront les trois situations que l'on vient d'évoquer.

1) Le concept d'"écart quadratique moyen" est (croyons-nous) d'origine française, mais l'expression est très lourde. L'anglais a choisi "standard-deviation". Après un essai malencontreux de "déviation standard", la terminologie statistique a, fort judicieusement, adopté "écart-type".

– L'anglais emploie indifféremment "variate" et (surtout) "random variable". Variate – mot nouveau à consonance française aussi bien qu'anglaise – a été quelquefois utilisé ; mais "variable aléatoire", utilisant des mots existants, est meilleur.

2) Pour désigner le "carré de l'écart quadratique moyen", l'anglais, langue efficace, a créé le mot "variance". Les statisticiens français l'ont adopté d'emblée : variance a une consonance parfaitement française ("invariance" était un mot français avant que "variance" ne soit homologuée par nos statisticiens).

– "Regression" constitue un exemple un peu ambigu des situations 1 et 2. Le sens usuel est le même en français et en anglais. L'emploi en statistique a été introduit par Galton ("Natural Inheritance" 1889) dans l'expression "regression to mediocrity" : "the sons of fathers who deviate x inches from the mean height of all fathers, deviate themselves from the mean height of all sons by less than x inches". Réserve étant faite sur l'expression : regression "to mediocrity", qui ne paraît pas en accord avec les données de Galton, reprises par Yule, on doit bien constater que le mot "regression" est maintenant utilisé dans le même sens par les statisticiens anglais et français, et qu'il n'est pas opportun de le remettre en question.

– L'exemple suivant mélange franchement les situations 1 et 2. On a signalé plus haut que "accuracy" doit se traduire par précision, et "precision" par fidélité : c'est la situation 1. Mais le concept de "précision (anglais) – fidélité (français)" se subdivise en deux sous concepts que les anglais ont baptisés "repeatability" et "reproducibility". Dans l'optique de la situation 2, la transposition en "répétabilité" (qualité de ce qui est répétable) et "reproductibilité" (qualité de ce qui est reproductible) n'offrirait aucune difficulté. A condition qu'on ait bien défini le "répétable" et le "reproductible"... voir plus loin.

– Un dernier exemple montre qu'il y a rarement des "cas désespérés", relevant de la situation 3. Faute d'équivalent existant, les informaticiens français ont longtemps utilisé les termes "hardware" et "software" – jusqu'au jour où l'un d'eux, bien inspiré, imagina "matériel" et "logiciel". Il est à souhaiter que ces deux mots nouveaux prennent solidement racine.

3) Un exemple de la 3^e situation est l'adoption telle quelle de l'anglais "probit". Selon le dictionnaire de termes statistiques de M.G. KENDALL et

W.R. BUCKLAND, ce mot a été introduit par BLISS en 1934, comme contraction de “probability-unit”. Plus de quarante ans d’utilisation (dans un domaine assez spécialisé de la statistique) justifient, selon nous, qu’on ne remette pas en cause l’emploi en français du mot probit. Il en est de même du terme “logit”, transformée $P' = \log \frac{P}{1 - P}$ de la probabilité définissant une loi logistique $P = [1 + e^{-(\alpha + \beta x)}]^{-1}$, d’où la relation linéaire $P' = -(\alpha + \beta x)$.

Mais il convient de prendre garde à de tels “condensés”, dont l’emploi semble se généraliser en langue anglo-saxonne. Nous ne voyons par exemple aucune utilité à ce que l’analyse de variance (“analysis of variance”) devienne l’“ANOVA”. Cette question des “mots condensés” et des sigles est suffisamment importante pour que nous lui consacrons ultérieurement une note distincte.

*
* *

Nous donnerons maintenant quelques exemples, parmi bien d’autres, de termes anglais qui n’ont pas actuellement d’équivalent français unanimement reconnu ; c’est sur ce problème que nous désirons attirer tout spécialement l’attention des lecteurs de la Revue de Statistique Appliquée qui ont bien voulu nous suivre jusqu’ici. Nous les citerons par ordre alphabétique, en les accompagnant de commentaires ou suggestions.

CONFOUNDING. – C’est un terme très familier aux spécialistes des plans d’expérience. Nous en rappelons ci-après la définition (anglaise) selon KENDALL et BUCKLAND (op. cit.).

“A device whereby, in large factorial experiments, the size of blocks is limited by sacrificing some of the independent comparisons relating to the higher-order interactions. These particular interactions may be deemed unimportant or of little practical consequence from a policy point of view.

The totality of possible treatment combinations is not replicated in each block but is divided amongst blocks in such a way that the main contrasts can be made within blocks but the others are not distinguishable from contrasts between blocks, with which they are said to be confounded.

More generally, when certain comparisons can be made only for treatments in combination and not for separate treatments, those treatment effects are said to be confounded. Confounding is often a deliberate feature of the design but may arise from inadvertent imperfections”.

La traduction littérale de “confounding” est “confusion”. Mais, dans le langage courant, ce substantif exprime l’état de ce qui est confus, en désordre (confusion mentale) ou même erroné (confusion de noms, de dates). Il n’est employé, dans le sens exprimé par le “confounding statistique”, que dans des expressions très particulières : confusion des pouvoirs, des peines, des patrimoines. Nous craignons que la traduction littérale n’entraîne, dans l’esprit des lecteurs, des confusions. . . et à notre connaissance elle n’a jamais été utilisée.

“Confounding” désigne à la fois le dispositif et l’action (ou opération) permettant de “confondre” — c’est-à-dire de rendre inséparable ou indiscernable — ou encore le résultat de cette action. Le mot français correspondant (à créer) pourrait être “inséparabilité” ou “indiscernabilité” ; mais il serait affreusement lourd, et exprimerait plutôt une “propriété”.

C’est la terminaison “age” qui exprime très souvent une action ou une opération :

“binage” : action de biner. “Calfeutrage” : action de calfeutrer. “Atterrisage” : action d’atterrir, ou opération permettant d’atterrir. “Arrondissage” : opération consistant à arrondir, etc. . .

Mais elle exprime aussi parfois,

– l’action et son résultat :

bricolage : action de bricoler (réparer tant bien que mal) et résultat de cette action (c’est du bricolage !)

– l’action et l’état qui en est la conséquence :

collage : action de coller et état de ce qui est collé (les collages de Matisse).

La terminaison “ion” est également fréquente (simulation, séparation, manipulation, . . .).

Puisque, pour traduire “confounding”, confusion doit être rejeté, pourquoi pas le néologisme “confondage” ?

FEED-BACK. — Pourquoi employer encore ce mot, alors qu’il existe en français le mot strictement équivalent de “rétroaction” ?

RANDOMISATION - TO RANDOMISE. — Les traducteurs ont jusqu’à maintenant buté sur le choix d’équivalents français. On dit couramment “randomiser”, “procéder à la randomisation”. . . Or, parmi les “ancêtres” de “random” on trouve l’ancien mot français “randon”, tombé en désuétude, d’où dérive notamment “randonnée”. Dire — et écrire — “randoniser”, “randonisation” serait un retour aux sources, et ne modifierait pas tellement les habitudes. . .

REPETITION - REPLICATION. — Nous empruntons encore les définitions ci-après à KENDALL et BUCKLAND.

“Répétition

A term denoting the execution of a statistical inquiry at different points in space or time, usually as part of a coordinated programme, as distinct from replication (q.v.).

Replication

The execution of an experiment or survey more than once so as to increase precision and to obtain a closer estimation of sampling error. Replication should be distinguished from repetition (q.v.) by the fact that replication of an experiment denotes repetition carried out at one place and, as far as possible, one period of time”.

Par un de ces illogismes dont aucune langue n'est exempte, à "replication" (. . . one place . . . one period of time) correspond, dans l'expression de la fidélité des mesures "repeatability", à "repetition" (different points in space or time) correspond "reproducibility".

De toutes façons la langue française ne dispose que d'un seul mot "répétition". Faut-il lui donner la seule signification de l'anglais "repetition", et trouver (ou créer) un mot correspondant à "replication" (avec ses dérivés "replicate", "duplication", "duplicate")? Nous ne le croyons pas. Cela serait inutile, le contexte, lors de l'emploi du mot répétition, devant être suffisamment clair pour éviter des erreurs d'interprétation. Et, dans la correspondance avec "répétabilité" et "reproductibilité", on se trouverait dans la situation illogique signalée plus haut.

RIDGE REGRESSION. — Il semble que ce terme ait tout d'abord signifié "régression avec contraintes" (imposées aux coefficients des "régresseurs"). D'où la suggestion qui a été faite de traduire par "régression bornée". Mais il semble aussi qu'il ait été ensuite utilisé pour désigner d'autres méthodes en analyse de régression. "Ridge regression" aurait donc un sens assez vague. Comment traduire — à moins de passer par l'intermédiaire du turc, voir ci-dessus — un terme qui n'a pas un sens bien défini, ou dont le sens initial s'élargit au gré des auteurs ?

SCREENING. — Selon KENDALL et BUCKLAND, "screening inspection" serait synonyme de "inspection à 100 %". Mais les auteurs français utilisent "screening" pour l'examen rapide d'un grand nombre de traitements, afin de ne retenir que les meilleurs, en vue d'une étude plus approfondie. Ce serait en somme un "passage au crible", ou "criblage".

SPLIT-PLOT. — Dans une expérience par la méthode des blocs, les variantes d'un premier traitement sont affectées aux parcelles ; chaque parcelle est elle-même divisée en sous-parcelles recevant les variantes d'un deuxième traitement. Le terme "blocs-subdivisés" semble convenir (de même qu'il y a "blocs complets", "blocs incomplets", etc. . .).

SPLIT-LEVEL. — Dans un essai destiné à déterminer la répétabilité d'une méthode, un même opérateur "répète" la mesure sur deux (ou plusieurs) échantillons identiques, ou aussi identiques que possible. On peut craindre que, lors des répétitions, l'opérateur ne soit, plus ou moins consciemment, influencé par le ou les résultats déjà obtenus. Afin d'éviter ce biais, on opère parfois d'une façon un peu différente. A l'opérateur sont présentés deux échantillons, non identiques, mais correspondant à des "niveaux" (de la grandeur étudiée) légèrement différents. Dans un document ISO⁽¹⁾ en cours de préparation, nous avons traduit "essai à niveaux subdivisés". (Il s'agit bien de subdiviser, chaque niveau correspondant à une division de l'ensemble des niveaux étudiés).

TREND. — Pourquoi certains auteurs emploient-ils ce mot, alors que "tendance" est rigoureusement équivalent ?

(1) International Standardization Organization.

Notre désir est que nombreux soient les lecteurs de cette Revue qui :

- nous fassent part de leurs observations sur le texte qui précède,
- nous signalent, avec commentaires et suggestions, les mots qui “posent problème” dans la traduction de l’anglais en français.