

ÉPILOGUE: STATISTIQUE ET MÉTRIQUE

Étienne ÉVRARD et Gerald PURNELLE
Université de Liège

Nous voudrions d'abord faire une remarque particulière, mais qui a son importance. Au cours de cette table ronde, nous avons beaucoup entendu parler de pourcentages. Nous pensons qu'une mise en garde est ici nécessaire. Non qu'il faille proscrire complètement ce mode de traitement: il permet de prendre une première idée d'un ensemble de données et il intervient dans un certain nombre de tests. Mais, employé sans discernement, il risque de conduire à des appréciations erronées. Le calcul des probabilités et, en particulier, la loi des grands nombres enseignent que, si une variable est distribuée aléatoirement, les valeurs obtenues au cours des épreuves ont une tendance à se grouper autour d'une valeur moyenne; de plus, les écarts par rapport à cette moyenne diminuent proportionnellement à mesure que le corpus grandit. Pour prendre un exemple banal, si on jette six fois un dé, on a peu de chances d'amener précisément une fois chacune des faces; en revanche, si on avait la patience de le jeter 60.000 fois, chacune des faces aurait un effectif très proche de 10.000. On voit ainsi que la dimension du corpus, c'est-à-dire, pour nous, la longueur du texte, joue un rôle. Ce rôle est complètement occulté par le calcul des pourcentages. Un phénomène observé dans deux corpus comparés peut présenter des pourcentages très différents sans que cela soit une indication significative d'un réel écart entre les deux corpus: il suffit, par exemple, que l'un d'eux soit petit et l'autre beaucoup plus abondant pour que la différence observée ne puisse raisonnablement être imputée qu'au seul hasard. Dépassant la simple observation des pourcentages, et appliquée à de tels recensements, la statistique permet de pondérer l'influence de la taille des corpus étudiés sur la proportion des effectifs de phénomènes attestés.

La métrique ancienne est un terrain idéal pour la statistique. Elle permet de compter divers niveaux de variables discrètes: syllabes, pieds, vers et même strophes, mots et situation des mots dans le vers, place des intermots, etc. Cependant, pour que ces données numériques se prêtent à

des calculs statistiques, encore faut-il que les variables choisies se réalisent sous diverses modalités, faisant ainsi apparaître des sous-ensembles que l'on peut comparer entre eux.

Si l'on choisit comme terrain d'investigation le mode de succession des pieds dans le vers, la plupart des mètres latins n'autorisent aucune recherche de ce type, dans la mesure où ils n'admettent aucune substitution et ne permettent donc pas la formation de sous-ensembles: c'est par exemple le cas de l'asclépiade majeur, dans lequel est écrite l'ode *carpe diem*. Même l'une ou l'autre substitution en petit nombre n'aide guère à constituer des sous-ensembles suffisamment caractérisés: que la quatrième syllabe du saphique, soit longue ou brève (ce qui est possible chez Catulle), la différenciation n'est pas suffisante pour qu'on s'en autorise à compter deux types distincts. Restent donc l'hexamètre dactylique ainsi que le sénaire iambique et le septénaire trochaïque.

Pour les deux derniers, il nous suffira sans doute de rappeler les travaux de L. Ceccarelli. Nous parlerons de préférence de l'hexamètre.

L'hexamètre dactylique admet des substitutions qui lui permettent de prendre 16 formes différentes (32 si l'on compte les spondaïques, à vrai dire très rares). Il devient donc possible de constituer des sous-ensembles en comptant les effectifs de ces 16 formes. À partir de ces données, une application statistique intéressante est facile à réaliser. Les effectifs des types d'hexamètre permettent de calculer le nombre de dactyles et de spondées à chacun des 4 (éventuellement 5) pieds à substitution et, de là, leur fréquence respective. On applique alors la loi des probabilités composées: par exemple, pour le vers formé d'une alternance dactyle, spondée, dactyle, spondée, on multiplie les fréquences du dactyle aux premier et troisième pieds par celles du spondée aux deuxième et quatrième. Ce produit indique la probabilité du types de vers choisi, en cas de répartition aléatoire, c'est-à-dire si l'emploi de tel pied à telle place n'a en rien été conditionné par la réalisation des pieds voisins. On multiplie cette probabilité par la longueur du poème considéré, ce qui donne un effectif théorique. Cette opération exécutée pour les 16 types hexamétriques fournit une distribution théorique, que l'on peut comparer à la distribution observée.

On notera que les fréquences sont, en fait, des pourcentages, mais que les effectifs théoriques, obtenus avec comme multiplicateur la longueur du texte considéré, font intervenir la dimension du corpus, ce qui élimine les critiques faites au début de cette note concernant les dangers des pourcentages.

La statistique offre le moyen de comparer une distribution observée à une distribution théorique; c'est le test de la variable χ^2 . Nous n'allons pas entrer dans le détail du calcul de cette variable; on le trouvera dans les manuels de statistique, et spécialement dans celui de Charles Muller, *Initiation aux méthodes de la statistique linguistique*, Paris, Champion, 1992 (1^{re} édition: 1973). Nous noterons seulement qu'une fois calculée la valeur de χ^2 , on peut en évaluer la signification au moyen de tables qui donnent la probabilité qu'un ensemble d'écarts au moins aussi important entre observation et prévision puisse apparaître aléatoirement.

C'est le moment de rappeler que, comme toujours en statistique, le test ne permet de rien affirmer avec certitude; il suggère seulement une interprétation des faits qu'il ne peut imposer. Soit un ensemble d'écarts qui n'a qu'une probabilité de 0,02 de correspondre à une situation aléatoire. On en conclura que, raisonnablement, l'hypothèse non aléatoire est de loin la plus vraisemblable et on est enclin à la choisir. Il n'en demeure pas moins vrai que, dans deux cas sur cent, c'est l'interprétation aléatoire qui doit être retenue. Au chercheur de choisir, en s'aidant d'autres indices pour diriger et étayer son choix.

Que nous apprennent de telles probabilités quant aux textes et à leur réception? On peut, à ce sujet, suggérer les réflexions suivantes. Soit un texte contenant un certain nombre de particularités qui le singularisent, mais d'une manière assez faible. Un lecteur (ou auditeur) attentif et très sensible va sans doute déceler l'anomalie, tandis qu'un autre, moins raffiné, ne va pas percevoir de différence. Un tel texte, d'après ce qui a été exposé précédemment, conduirait sans doute à un χ^2 correspondant à une probabilité aléatoire d'environ 0,50 ou 0,60, ce qui signifie qu'il a des chances à peu près égales d'être purement aléatoire ou d'être significatif. Cela correspondrait bien aux réactions des deux témoins considérés ci-dessus: l'un des deux serait plus sensible aux aspects significatifs, tandis que l'autre ne dépasserait pas l'aspect aléatoire. Soit, en revanche, un texte où les singularités seraient en quantité surabondante. Les deux témoins vont certainement, bien qu'à des degrés variables, sentir l'un et l'autre le caractère significatif du texte quant à sa métrique. Corrélativement, la probabilité correspondant au χ^2 devrait, dans un tel cas, se réduire à quelques centièmes. En d'autres termes, il ne semble pas exagéré de suggérer que la probabilité dégagée par le test correspond grosso modo à l'impression de singularité ressentie par un lecteur, en fonction de son degré de sensibilité.

Soumis à un tel test, certains poèmes hexamétriques latins se révèlent tout à fait aléatoires; cela signifie que leurs auteurs n'ont tenté, ni

consciemment ni inconsciemment, d'agir sur le mode de succession des pieds, ou que, l'ayant tenté, il n'y ont pas réussi. D'autres, en revanche, ont une disposition des pieds franchement significative; il restera alors à examiner quels sont les types hexamétriques responsables de cette particularité, à déterminer ce qu'ils ont en commun et à chercher à quel dessein pourrait répondre une telle recherche, qu'elle soit consciente ou non. Il arrive même que, dans des poèmes d'une certaine longueur, on détecte un manque d'homogénéité entre diverses parties; on sera alors amené à déterminer dans quelle mesure la métrique peut contribuer à la structuration de l'œuvre. Entre ces extrêmes existent évidemment de multiples intermédiaires qu'il faudra apprécier en fonction des circonstances. Il y a là une mine d'investigations passionnantes qui, menées avec prudence et rigueur, sont de nature à affiner notre connaissance de la poésie latine.

Jusqu'ici, nous n'avons envisagé que la succession des pieds, telle qu'elle se réalise dans les vers. On se souvient que, de ce point de vue, les recherches n'étaient possibles que pour les mètres admettant des substitutions. Mais il est d'autres objets possibles, et que l'on peut étudier dans toute espèce de vers. Nous en citerons quelques-uns: la place des intermots, problème lié à une étude rigoureuse des césures; la correspondance ou la discordance entre mots et pieds, et, subsidiairement, la coïncidence ou la non-coïncidence entre l'ictus métrique et l'accent verbal; la répartition des formes dans le vers et, en particulier, pour les formes très fréquentes dans un poème, le fait qu'elles occupent indifféremment les diverses places qu'autorise leur prosodie ou qu'elles se massent en majorité à certaines positions. Chaque fois, la méthode est la même: on part de l'hypothèse nulle, où le phénomène n'obéit qu'aux aléas du hasard, on détermine les effectifs qui en résulteraient et on les compare aux effectifs observés.

Nous n'en dirons pas plus. Notre intention était seulement de montrer comment le domaine de la métrique est susceptible d'un traitement statistique efficace et, corollairement, qu'il est préférable de ne pas s'y contenter de simples dénombrements et d'impressions subjectives.