

Présentation d'un programme automatique d'apparat critique positif

par

Ghislaine VIRE

Université Libre de Bruxelles - BELGIQUE

891

Les philologues qui appliquent les méthodes automatiques de traitement des données au domaine de la critique textuelle et de l'édition de texte n'ont pas tous la même approche méthodologique des problèmes spécifiques à cette branche de la philologie. C'est la raison pour laquelle, avant de décrire le programme d'apparat critique positif qui fait l'objet de mon exposé, je voudrais situer très rapidement ma démarche sur le plan de la méthode.

L'utilisation de procédures automatiques m'a été dictée par l'ampleur du sujet que je me proposais d'étudier, à savoir la tradition manuscrite d'un texte de l'antiquité latine qui ne comprend pas moins d'une centaine de manuscrits. Il s'agit d'une oeuvre mineure de la littérature latine, le *de astronomia* d'Hygin, composé dans les dernières années du 1er siècle av. J.C., mais qui doit, à son inspiration en grande partie mythologique et au caractère pragmatique de ses considérations astronomiques, d'avoir connu un franc succès pendant tout le moyen âge et, plus particulièrement, durant les IXème, Xème et XIème siècles.

Ma démarche est donc exclusivement pratique, dans la mesure où elle vise, avant tout, à partir des données fournies par la collation complète non de la totalité des manuscrits, mais des quarante plus anciens d'entre eux, tous antérieurs à l'époque humanistique, à localiser, dans le temps et dans l'espace, les points de repère les plus importants de l'histoire du texte, à reconstituer l'enchaînement des copies dont rend compte l'état actuel de notre documentation et, enfin, à établir une édition dont le texte reflète aussi fidèlement que possible ce qu'a dû être le plus proche commun ancêtre de la tradition (1).

Le pragmatisme délibéré de cette démarche et l'exigence qui l'accompagne de fonder le classement automatisé des manuscrits sur des dépouillements exhaustifs plutôt que sur quelques collections de variantes, comme nombre de philologues se résignent encore trop souvent à le faire, se doublent de la volonté de recourir, dans toute la mesure du possible, pour traiter les données, à des programmes généraux existants, comme les programmes d'index, de concordance ou de tri. Aujourd'hui plus encore qu'hier il me semble hors de question que celui qui se sert de l'informatique dans un domaine, comme celui des sciences humaines, où les problèmes à résoudre peuvent souvent être ramenés à une succession de traitements relativement simples, s'offre le luxe de faire élaborer, ou d'élaborer lui-même, au prix d'un long travail, des programmes dont l'utilisation ne peut dépasser le cadre d'un sujet particulier de recherche. Adapter, moyennant la création de quelques routines supplémentaires, des programmes existants à un domaine d'étude auquel ils ne s'appliquent pas encore ne peut être que bénéfique pour le chercheur, dans la mesure même où celui-ci se voit obligé de repenser les méthodes de travail auxquelles il a été formé.

La marche à suivre que j'ai adoptée a pour objectif de regrouper les quarante manuscrits dont je dispose en fonction des variantes communes qu'ils partagent. Pour y parvenir, j'ai eu recours à un traitement automatisé dont les étapes successives visent à caractériser davantage, puis à restructurer les données issues des collations, pour déboucher finalement sur la liste hiérarchisée de tous les groupes de deux à trente manuscrits que définissent les variantes de toutes sortes transmises par la tradition.

Pour expérimentale qu'elle soit, ma démarche n'en est pas moins générale, je crois, étant donné qu'elle est fondée sur des principes unanimement admis en critique textuelle et mise en oeuvre à l'aide de procédures qui ont un champ d'application très vaste. Je compte d'ailleurs la tester, dans un avenir relativement proche, sur un autre texte de l'antiquité latine, le *de die natali* de Censorin.

J'en viens, au terme de cette introduction, au programme d'apparat critique positif. Le but que poursuit le philologue, une fois qu'il est parvenu à retracer l'histoire du texte qu'il étudie, consiste à dresser la liste de tous les lieux variants attestés sinon par l'ensemble de la tradition, du moins par ceux de ses représentants qui méritent d'être pris en compte lors de l'établissement de l'édition. Le programme d'apparat critique positif a, de la même façon, été conçu comme le prolongement logique des procédures utilisées pour classer les manuscrits, puisqu'il vise à synthétiser les données des deux fichiers qui ont servi de base au regroupement des leçons communes.

Le premier de ces deux fichiers, le fichier de référence, contient la transcription intégrale de celui des manuscrits qui a été choisi comme point de comparaison lors de la collation des autres copies. Ce manuscrit a été choisi de manière arbitraire, en fonction de son ancienneté, de sa lisibilité et de son accessibilité. Il ne transmet donc pas nécessairement l'état du texte le meilleur et le plus efficace, lorsqu'il s'agit d'établir le fichier des variantes, ni, à fortiori, celui qui a des chances d'être le plus adéquat pour établir l'édition. L'un des avantages du programme ACAP est précisément de permettre, une fois le classement des variantes établi, de substituer au manuscrit de référence initialement choisi un autre exemplaire dont le texte s'est avéré être de meilleure qualité à l'issue du traitement. L'enregistrement du manuscrit de référence n'a nécessité aucune manipulation particulière si ce n'est la normalisation de l'orthographe et l'uniformisation des noms de nombre, tous transcrits en chiffres arabes. Traité à l'aide du programme d'index (GITEX), dont une des options permet de numéroter séquentiellement les mots d'un texte, le manuscrit de référence a été transformé en un fichier de référence dont chacun des éléments est identifié à l'aide d'une référence non ambiguë, qui comprend le numéro du livre, le numéro du chapitre et le numéro du mot au sein du chapitre (2).

Si le fichier de référence comprend les quelque vingt-et-un mille mots que totalise le *de astronomia* d'Hygin, le second fichier, celui des variantes, compte plus de soixante mille éléments, puisqu'il contient toutes les leçons qui diffèrent de celles du texte de référence, exception faite des divergences orthographiques, et qu'au surplus il renferme les corrections et annotations marginales ou interlinéaires dues à l'intervention des copistes médiévaux. Le fichier des variantes est plus que le simple appareil critique négatif établi à partir du texte de référence, étant donné qu'y ont été insérées des catégories supplémentaires de leçons indispensables au classement : à titre d'exemple, je citerai les variantes qui sont conservées par le manuscrit de référence et par un petit nombre de copies — moins de dix — dans des endroits du texte où les autres exemplaires ont une forme banalisée. Il va de soi que les données du fichier des variantes, enregistrées au fur et à mesure de la progression des collations, ont fait l'objet de la même normalisation que celles du fichier de référence (2).

Le programme d'apparat critique positif est destiné à fournir au philologue l'instrument de travail dont il a besoin pour établir le texte. C'est dire qu'il regroupe, pour chaque lieu variant qui a fait l'objet d'altérations au cours des transcriptions successives, tous les éléments fournis par la tradition, compte tenu de l'état fragmentaire de certaines des copies et du fait qu'une leçon peut, dans certains exemplaires, correspondre à deux ou plusieurs mots. Le principe de fonctionnement du programme est très simple : partant du fichier des variantes établi d'après l'ordre d'apparition des lieux variants dans l'oeuvre, le programme identifie toutes les leçons qui portent la même référence, puis il recherche le mot correspondant dans le fichier de référence et fait suivre automatiquement celui-ci des manuscrits dont le sigle ne figure pas dans le fichier des variantes. Pour fonctionner correctement, le schéma qui vient d'être décrit très succinctement appelle quelques remarques. Pour que le fichier de référence et celui des variantes soient compatibles, il faut éliminer du fichier de référence les catégories particulières de leçons qui y avaient été ajoutées dans le but de rendre le classement des manuscrits

parfaitement correct. Il s'agit aussi d'enregistrer les lacunes des exemplaires incomplets, comme s'il s'agissait d'omissions longues, sans quoi le programme restituera le sigle de ces manuscrits à la suite de la leçon du texte de référence, dans les passages qu'ils n'ont pas. Enfin, pour faciliter la tâche du philologue, il convient de substituer l'abréviation CODD, usuelle pour *codices*, dans tous les cas où, à une forme individuelle du texte de référence, s'oppose la leçon de l'ensemble de la tradition.

D'autre part, des traitements particuliers doivent être réservés aux corrections et annotations ainsi qu'aux variantes qui englobent plus d'un mot. En ce qui concerne les corrections et les annotations la difficulté provient de ce que, lorsqu'une copie transmet à la fois une correction et une leçon différente de celle du texte de référence, son sigle apparaît deux fois dans le fichier des variantes ; de même, l'exemplaire qui conserve la même leçon que le texte de référence, mais a été corrigé ultérieurement figure, lui aussi, dans le fichier des variantes. La solution consiste à isoler, dès le début de l'examen d'un lieu variant, les formes qui portent la mention correction (COR) ou annotation (SV ou MG, selon qu'il s'agit d'une annotation interlinéaire ou marginale) et de ne pas tenir compte de ces données en cours de traitement mais de les réinsérer à la fin de l'unité critique lors de la mise en page. Le second cas particulier concerne les omissions longues, les lacunes des manuscrits fragmentaires qui leur ont été assimilées, ainsi que les variantes, comme les inversions, qui portent sur plus d'un mot et qui, pour des raisons d'ordre philologique, ont été considérées comme formant un seul lieu variant plutôt que subdivisées en autant de lieux variants qu'elles comprennent de mots. Des complications surviennent dans les cas où coexistent, au sein du fichier des variantes, une leçon portant sur un intervalle de plusieurs mots et une ou plusieurs autres altérations affectant un seul mot de cet intervalle. Pour surmonter cette difficulté, il suffit de conserver dans un fichier intermédiaire les références des premier et dernier mots des lieux variants longs, ainsi que les sigles des manuscrits qui les conservent.

La mise en page des résultats a été établie de manière telle que l'utilisateur soit en mesure d'identifier très rapidement toutes les données qui se rapportent à un même lieu variant. Par ailleurs, les mots qui n'ont subi aucune altération lors de la transmission sont également imprimés, dans le but d'assurer la parfaite continuité du texte à celui qui dépouille l'apparat critique et de lui faciliter, du fait même, la compréhension des différentes leçons attestées. A ces avantages liés à la présentation s'en ajoute un autre qui a trait à l'utilisation du mode conversationnel, qui permet au philologue de négliger, lors de l'impression, les leçons qui ne présentent guère d'intérêt à ses yeux, pour ne retenir que celles qu'il compte faire figurer dans son édition.

Par manque de temps, je n'ai encore utilisé le programme d'apparat critique positif que pour des chapitres isolés de l'oeuvre. Il va sans dire que, très prochainement, je compte l'appliquer à la totalité du texte et je me propose, à la faveur de ce traitement, de tester quelques hypothèses de travail qui me semblent intéressantes du point de vue méthodologique et que je voudrais brièvement présenter pour terminer. L'une des questions qui restent aujourd'hui encore sans réponse concerne le choix du manuscrit de référence. On admet généralement a priori que l'exemplaire qui sert de base aux collations peut être choisi de façon arbitraire, dès le moment où le fichier des variantes établi à partir de lui ne sert qu'à dégager un stemma non orienté des manuscrits disponibles. A partir du moment où l'on s'impose des dépouillements exhaustifs et que l'on dispose d'un programme d'apparat critique positif capable de synthétiser l'ensemble des données de la tradition, plus aucun obstacle matériel ne s'oppose à ce que les mêmes procédures de classement soient utilisées à plusieurs reprises successives, en prenant lors de chaque traitement un manuscrit différent comme texte de référence. Grâce à des programmes semblables à celui que je viens de décrire, il ne faut plus, pour effectuer ce travail, ni compléter ni transformer manuellement les deux fichiers, ce qui représente un gain de temps considérable et élimine une source importante d'erreurs.

Une autre direction de recherche consiste à tenter de définir des critères qui permettent de choisir rapidement le manuscrit de référence susceptible de fournir le fichier des variantes le plus performant, c'est-à-dire celui où ne figurent que des leçons utiles pour le classement. L'utilisation du programme d'apparat critique positif devrait permettre, dans ce cas-ci aussi, de changer de texte de référence une ou plusieurs fois en cours de travail, ce qui autoriserait le philologue à mesurer l'efficacité et les effets de tels changements.

En guise de conclusion, je rappellerai simplement qu'en plus de son utilité immédiate en critique textuelle, le programme automatique d'apparat critique positif constitue pour le philologue un nouvel instrument d'analyse de certains des problèmes méthodologiques que posent la collation et le classement des manuscrits. N'est-ce pas précisément ou ne devrait-ce pas être une des fonctions de l'informatique que de susciter chez l'utilisateur d'autres approches des problèmes que celles auxquelles l'ont accoutumé les traditions de sa discipline ?

NOTES

- (1) Les conclusions de mes recherches portant sur la transmission du texte du *de astronomia* jusqu'au XIII^e siècle doivent être publiées dans le prochain numéro de la *Revue d'Histoire des Textes*.
- (2) Les principes méthodologiques qui ont présidé à l'élaboration du fichier de référence et de celui des variantes ont fait l'objet d'un exposé succinct au colloque : *La pratique des Ordinateurs dans la critique des textes*, à Paris en mars 1978 (cf. *Colloques Internationaux du C.N.R.S.*, n° 579, Paris, 1979, pp. 51-60).
La méthode que j'ai suivie pour regrouper les variantes sera explicitée très prochainement dans une publication détaillée.