HAL open science

Transcrire des écrits scolaires : entre philologie et génétique textuelle

Pierre-Yves Testenoire

To cite this version:

Pierre-Yves Testenoire. Transcrire des écrits scolaires : entre philologie et génétique textuelle. Corpus, 2017, Spécificités et contraintes des grands corpus de textes scolaires : problèmes de transcription, d'annotation et de traitement, 16, p. 87-109. halshs-02092110

HAL Id: halshs-02092110

https://shs.hal.science/halshs-02092110v1

Submitted on 7 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Transcrire des écrits scolaires : entre philologie et génétique textuelle

Pierre-Yves TESTENOIRE
Université Paris-Sorbonne, ESPE de Paris /
UMR Histoire des Théories Linguistiques

La constitution d'un corpus d'écrit scolaire implique, entre autres opérations, un acte qui ne paraît pas d'une technicité extrême : celui de transcrire. La simplicité de la définition que le Littré en donne – « transcrire : copier un écrit » – ne rend cependant pas compte de la complexité des problèmes inhérents à cette opération. On se propose de réfléchir aux problèmes méthodologiques posés par la transcription d'un écrit d'élève comme de tout autre manuscrit moderne et à l'adéquation à cet objet spécifique des outils et des modèles théoriques disponibles. On le fera à partir des deux champs disciplinaires qui ont placé le manuscrit au centre de leurs investigations : la philologie et la génétique textuelle. On examinera les caractéristiques épistémologiques de ces deux approches pour réfléchir aux modalités de la transcription envisagées dans ces deux cadres.

1. Le manuscrit moderne, entre philologie et génétique textuelle

La génétique textuelle s'est développée dans les années 70 sur la revendication d'une rupture par rapport à l'approche philologique des manuscrits. L'évaluation de la nature de la différence entre génétique textuelle et philologie dépend de la conception que l'on se fait de cette dernière. Celle-ci a reçu, dans les classifications des sciences proposées depuis le XIXe siècle, un rôle plus ou moins ambitieux : science englobant toutes les connaissances du passé (Wolf, Böckh), science de la critique (Renan), science du langage humain et de ses productions (Auebarch), science historique (Wilamowitz, Usener), voire simple commentaire au bas d'une littérature (Saussure)… On partira ici d'une conception minimale de la « philologie » entendue comme « science du texte », définition également revendiquée à l'occasion par la

génétique textuelle mais dans un sens différent. L'affirmation développée par les pionniers de la génétique du texte selon laquelle elle constituerait un changement de paradigme repose sur quatre arguments principaux.

En premier lieu, philologie et génétique constitueraient deux disciplines distinctes car elles auraient des objets scientifiques différents. L'objet de la philologie est le « texte » que l'analyse des manuscrits sert à restituer. Pour la philologie classique et médiévale, c'est le texte originel perdu, l'*Urtext*, que le philologue cherche à reconstituer via les manuscrits copiés par des scribes. Appliquée aux manuscrits de création, la perspective philologique traite le manuscrit comme une source du texte définitif dont elle relève au besoin les variantes. La génétique s'attache, quant à elle, à la genèse d'un texte, autrement dit à son processus de création. Elle n'accorde pas *a priori* de supériorité axiologique au texte publié plutôt qu'au manuscrit, à ce qui est conservé plutôt qu'à ce qui est biffé : le manuscrit est abordé sans perspective téléologique.

Cette différence d'objet est rattachée, en second lieu, dans les textes des généticiens à des ancrages théoriques distincts. La génétique se développe en France dans un contexte spécifique : celui de la fin du structuralisme. Le tournant des années 60 et 70 voit le développement des théories du texte (Kristeva, Barthes, Todorov, Riffatterre…) dont s'affranchit progressivement la perspective génétique. L'observation des manuscrits modernes conduit à remettre en cause la conception structurale du texte et sa clôture sémiotique, spatiale, temporelle. Cette remise en cause passe par la mise en place de la notion d'avant-texte (Bellemin-Noël 1972) puis par la proposition d'une « conception partitive du texte » (Lebrave 1986) : le manuscrit ne serait pas *un* texte mais *du* texte. En 1985, dans un article au titre provocateur « Le texte n'existe pas », Louis Hay propose la liquidation du concept devenu inopérant pour les études de genèse. Pour réintroduire la dimension productive et temporelle dans l'analyse de l'écrit, la génétique trouve ses modèles théoriques dans la linguistique de l'énonciation, en réfléchissant aux spécificités de l'énonciation écrite (v. Fuchs *et alii* 1982). La philologie, quant à elle, reste attachée à un modèle théorique où le texte est un concept qui garde sa pertinence comme entité stable, invariante et synchronique.

Troisième différence majeure entre linguistique et philologie : ils porteraient sur des objets empiriques différents. Du

point de vue de constitution des deux champs, cette différence est première par rapport au cadre théorique. Ce sont, en effet, des circonstances liées à la réunion de chercheurs autour des manuscrits modernes, au sein du Centre d'Analyse des Manuscrits fondé en 1975 par Louis Hay et devenu le laboratoire ITEM en 1982, qui ont conduit à développer des pratiques et des théories propres à cet objet. Ce qui fonde la distinction entre les objets d'études de la philologie et de la génétique est moins l'opposition entre manuscrits anciens d'un côté et manuscrits modernes de l'autre que l'opposition entre manuscrits de transmission et manuscrits de création. Les manuscrits de création étant exceptionnels avant le XVIIe siècle, les deux clivages ont cependant tendance à se superposer.

Des différences d'ordre méthodologiques, enfin, séparent les deux champs. Conformément à son cadre théorique et à ses objectifs, la critique génétique a développé une méthodologie spécifique, reprenant et adaptant certains gestes philologiques (déchiffrer, transcrire), en abandonnant d'autres (pratiquer l'*emendatio*, établir un *stemma*) et codifiant de nouvelles opérations (constituer un dossier de genèse, par exemple). De même, elle a crée ses propres notions (avant-texte, opérations d'écriture, campagne d'écriture…) et en a rejeté ou redéfini d'autres héritées de la tradition philologique (texte, variante, brouillon, etc.).

C'est sur la base de cette série d'arguments épistémologiques que Jean-Louis Lebrave pouvait affirmer en 1992, dans le premier numéro de la revue *Genesis* que la génétique textuelle était une « discipline nouvelle » et non « un avatar moderne de la philologie ». Vingt-cinq ans plus tard, le débat sur la relation entre philologie et génétique semble plus apaisé, et les positions moins tranchées. Le même Lebrave écrit ainsi, en 2009, qu'« en tant qu'elle est une poétique des transitions entre états, la critique génétique est bien la fille de la philologie, ou plutôt une proche parente de la « nouvelle philologie »[1], mais aussi de la *variantistica* de Contini ».

[1] Signalons qu'un des promoteurs de cette « nouvelle philologie » à laquelle pense Lebrave, Bernard Cerquiglini, est un compagnon de route de la génétique textuelle. Dans le sillage de son analyse critique de la philologie médiévale développée dans *Éloge de la variante*, il a collaboré aux travaux des généticiens : v. par exemple, Cerquiglini 1998 et 2010.

Cette relativisation de la différence entre philologie et génétique est en partie liée à l'évolution des deux champs. La philologie n'est plus cette science positiviste éreintée par le structuralisme : les nouvelles technologies et le développement des *digital humanities* lui ont profité, notamment du point de vue des politiques de financement de recherche[2]. La génétique, quant à elle, a depuis les années 70 élargi son champ aux textes non littéraires – philosophiques, scientifiques, scolaires – et aux créations non uniquement textuelles – musicales, graphiques, arts vivants, architecture, etc. (v. Biasi et Herschberg Pierrot 2015). La contestation du domaine de la philologie n'a plus, dans cette configuration, la même actualité.

Avec l'effacement de cet enjeu, des zones de rencontre entre les deux disciplines sont apparues. Ainsi que l'a fait remarquer Michel Espagne (2010), la génétique textuelle s'est développée dans un contexte français où la tradition philologique était faible et sans grande ambition théorique. Or, il est d'autres contextes où des réflexions communes à celles développées par la génétique ont été menées dans le cadre philologique, par exemple en Italie avec la *filologia d'autore* et la *critica delle varianti* (Segre 1995, 2010). On peut également identifier, par delà les différences revendiquées, des attitudes communes au philologue et au généticien face au manuscrit : ainsi en va-t-il de la *lectio difficilior*. Ce principe philologique, qui veut qu'entre deux leçons de manuscrits la plus difficile, celle dont l'interprétation est la moins aisée, aura tendance à être la plus authentique en vertu de la simplification inhérente à l'activité de copie, est d'une certaine façon reprise par le généticien qui préfèrera aussi le manuscrit difficile à lire – le brouillon extrêmement raturé plutôt que sa mise au net –, seul à même de renseigner sur la genèse.

Au regard du temps long de l'histoire des sciences, la génétique textuelle relève donc évidemment de la philologie conçue comme science du texte depuis la période alexandrine. Cela n'exclut pas qu'elle ait forgé des méthodes et des conceptions nouvelles par rapport à celles mises en place par la philologie moderne au début du XIXe siècle pour l'étude des textes antiques. La coexistence d'éléments de continuité et d'innova-

[2] V. aussi Duval (2007) qui analyse les raisons épistémologiques de ce « retour en grâce » de la philologie.

tion est patente dans le cas d'une des opérations que pratique aussi bien le philologue que le généticien : la transcription d'un manuscrit.

2. Transcrire un manuscrit : objectifs, choix, contraintes, modèles

Toute transcription de manuscrits, indépendamment du cadre théorique dans lequel elle s'inscrit, est soumise à la tension résultant de la poursuite de deux objectifs contradictoires : un objectif de *vi-lisibilité* et un objectif de *lisibilité*.

La *vi-lisibilité* est un néologisme forgé par Jacques Anis (1983) pour désigner les propriétés visuelles d'un texte qui participent de sa signifiance. L'exigence de *vi-lisibilité* est celle qui tend à restituer la disposition et les caractéristiques visuelles de l'écrit manuscrit. L'exigence de *lisibilité* tend inversement à présenter au lecteur le texte le plus clair et accessible possible, c'est-à-dire conforme aux conventions typographiques usuelles. C'est à cette double fidélité qu'est assigné le transcripteur dans son opération de transcodage : fidélité au système sémiotique de départ (l'écriture manuscrite) et fidélité au système sémiotique d'arrivée (l'écriture typographique)[3]. L'exigence de *vi-lisibilité* procède par reproduction mimétique des caractéristiques sémiotiques du manuscrit (soulignements, ratures, disposition spatiale de l'écrit…) quand la priorité accordée à la *lisibilité* favorise la linéarisation de la transcription et l'usage de signes critiques et de conventions éditoriales.

Les choix de transcription sont une négociation entre ces deux exigences contradictoires. Ils sont déterminés, pour chaque corpus manuscrit, par trois paramètres principaux :

- **Ce pour quoi on transcrit**. Les objectifs assignés à la transcription sous-tendent les choix effectués. De l'exploitation envisagée du corpus dépendent les paramètres de l'écrit que l'on jugera pertinent de retenir. Si la transcription s'inscrit dans une perspective éditoriale, le lectorat visé (grand public, spécialisé, etc.) influe sur le degré de technicité des conventions éditoriales. Si la transcrip-

[3] Pour une réflexion sémiotique sur l'écriture typographique et manuscrite, v. Arabyan 2012 : 11-103.

tion est destinée au chercheur (description du manuscrit par le généticien, constitution d'un corpus informatisé), les objectifs de recherche sont également implicités par la transcription. Ainsi dans le cas de la constitution du corpus *écriscol*, l'intérêt pour les erreurs orthographiques et les problèmes de segmentation des unités présents dans les écrits d'élèves a justifié la transcription telles quelles des formes non normées.

- **Ce que l'on transcrit**. Le manuscrit moderne se caractérise par son caractère protéiforme. La plasticité permise par l'écriture manuscrite et son cantonnement progressif dans la sphère privée ont conduit certains scripteurs à investir le manuscrit comme un espace de liberté et d'affranchissement des normes et des contraintes de l'écrit imprimé. Dès lors, une multitude de paramètres contribue à faire varier les manuscrits : graphie plus ou moins lisible, spatialisation plus ou moins linéaire de l'écrit, affranchissement plus ou moins poussé des normes linguistiques et typographiques, présence ou non de signes non verbaux, d'abréviations, de ratures… À cette variabilité s'ajoutent des variations fonctionnelles : les manuscrits d'un même auteur peuvent présenter de grandes différences, selon qu'il s'agit d'un texte écrit pour soi-même ou pour autrui, de notes, d'un brouillon, d'une mise au net… Une telle diversité explique en partie l'absence d'harmonisation des conventions de transcription, régulièrement déplorée. La nécessité d'adapter les conventions de transcription aux problèmes posés par chaque corpus manuscrit prime bien souvent dans les faits sur le souhait d'harmonisation.
- **Ce avec quoi on transcrit**. Les technologies mobilisées pour la transcription constituent le troisième paramètre majeur rentrant en ligne de compte dans les choix effectués. Ce ne sont pas seulement les techniques et les logiciels de transcription qui jouent un rôle déterminant, mais aussi le support final. Le support papier ou numérique, la possibilité d'intégrer des images, les ressources ouvertes par l'hypertexte sont autant de variables qui délimitent et orientent les choix de transcription.

Si ces trois grandes variables déterminent les choix du transcripteur, les modalités de transcriptions ne sont pas illimitées. La philologie classique et médiévale distingue traditionnellement trois modes d'éditions, impliquant ou non des gestes de transcription spécifiques :

- **L'édition critique.** Elle propose la reconstitution d'un texte original par l'examen critique et comparé des différents manuscrits de sa tradition. Elle passe soit par la transcription d'un seul manuscrit considéré comme un bon témoin (bédiérisme) soit par la sélection pour chaque passage, parmi les manuscrits, de la leçon analysée comme authentique (méthode de la philologie classique dite lachmanienne). Dans les deux cas, la transcription est inséparable d'une critique textuelle qui s'appuie sur une méthodologie largement manualisée. L'éditeur pratique la conjecture : il restitue une leçon qu'il juge authentique si elle ne lui paraît pas attestée par le ou les manuscrits. La description codicologique des manuscrits et leurs variantes sont données dans l'apparat critique ou dans une section ecdotique.
- **L'édition diplomatique**. Ses principes remontent au *De Re Diplomatica* de Mabillon (1681). Conçue pour la restitution exacte des actes juridiques, la transcription diplomatique a ensuite été utilisée par la philologie moderne pour les manuscrits d'auteurs et pour les manuscrits de copie des textes médiévaux. Elle consiste en « un relevé archéologique des textes, tels qu'ils sont transcrits par les manuscrits existants » (Masai 1950 : 185). La transcription diplomatique ne s'attache qu'à un seul manuscrit dont les caractéristiques matérielles et graphiques sont rendues grâce à des conventions éditoriales.
- **L'édition en fac-similé.** Elle consiste en la reproduction photographique de la page manuscrite. Comme elle ne fait pas intervenir l'opération de transcription, elle ne nous intéressera pas ici.

Ces trois modèles sont trois voies de reconstitution du « texte » original ; ils correspondent à différents degrés d'interventions de l'éditeur sur la leçon de chaque manuscrit. La génétique textuelle a redéfini les enjeux de la transcription en fonction ses objectifs propres : la reconstitution du processus d'écriture.

C'est le degré de chronologisation des opérations d'écriture qui répartit les trois types de transcription auquel elle a recours :

- **La transcription diplomatique**. Elle vise, comme son ancêtre philologique, la reproduction exacte du manuscrit. La différence majeure est que cette exactitude n'est pas obtenue par des conventions éditoriales, mais par une reproduction mimétique. La transcription diplomatique génétique peut donc être définie comme « la reproduction dactylographique d'un manuscrit qui respecte fidèlement la topographie des signifiants graphiques dans l'espace : chaque unité écrite figure à la même place de la page que sur l'original » (Grésillon 1994 : 246).
- **La transcription linéarisée**. Elle consiste à transcrire tous les signifiants graphiques présents sur le manuscrit dans une ligne continue de signes. Ce qui est hors ligne est linéarisé à l'aide d'un codage typographique. La succession des opérations est notée par la consécution des signes graphiques sur la ligne et grâce à des conventions : généralement, < xxx > pour noter un ajout, ~~xxx~~ ou [xxx] pour une suppression, et leurs combinaisons pour coder un remplacement ou un déplacement. Ce type de transcription, dont le résultat est le plus proche des canons de l'imprimé, a la faveur des éditeurs ; il présente aussi des avantages techniques pour les traitements automatiques. La transcription linéarisée pose néanmoins problème en ce qu'elle propose une lecture univectorielle de l'ensemble des traces graphiques laissés sur un manuscrit. Comme l'écrit Lebrave (1990 : 145) elle « introduit en quelque sorte clandestinement des éléments [d'analyse] dans une transcription qui se veut transparente ».
- **La transcription chronologique.** Elle a été développée par certains généticiens pour pallier les limites des deux types de transcriptions précédents. Jean-Louis Lebrave en particulier, a proposé des exemples de transcription chronologiques de manuscrits de Heine (1984) et de Proust (1990). Conçue comme une visualisation de la genèse, la transcription chronologique repose sur le principe de la substitution génétique auxquelles sont ramenées les quatre opérations d'écriture fondamentales (ajout, suppression, remplacement et déplacement). Elle consiste

donc à reconstituer les différents états de l'écrit entre lesquels est intervenue une substitution génétique. Les étapes transcrites ne correspondent donc pas nécessairement aux segments attestés tels quels sur le manuscrit. Ce type de transcription, insiste Lebrave (1990), est destiné à des analyse micro-génétiques et peut être utile dans la perspective d'un traitement automatique des substitutions. Il reste cependant peu utilisé, même par les généticiens.

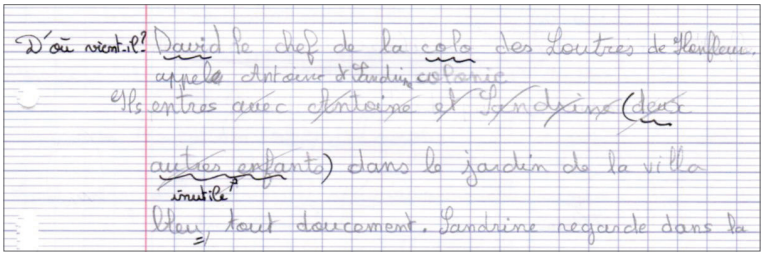
Notons que la génétique textuelle n'a pas apporté d'innovations majeures dans l'opération de transcrire. Seule la transcription chronologique, qui reste peu pratiquée, peut lui être attribuée : pour le reste, elle a repris les principales conventions de la tradition philologique (les chevrons par exemple dont l'histoire remonte à la philologie alexandrine) en affinant son modèle au contact de la complexité de certains manuscrits modernes.

Ces trois modèles de transcription ont parfois été classés en fonction du degré d'interprétation qu'ils supposent : la diplomatique serait « la plus objective », la chronologique « la plus interprétative », la linéarisée occuperait une position intermédiaire (v. par ex. Crasson et Fekete 2004). Si ce classement peut être accepté du point du vue du transcripteur, il s'estompe dès que l'on se place du point de vue de la réception. Dans la perspective génétique, en effet, la transcription est un outil pour interpréter un processus d'écriture. Cette interprétation consiste à convertir des données visuelles (disposées sur le manuscrit) en données temporelles (des opérations d'écriture). Les différents types de transcription se distinguent donc moins parce qu'ils sont « plus ou moins interprétatifs » que dans la façon dont ils distribuent les rôles dans cette opération : dans le cas de la diplomatique, la charge de la conversion des données visuelles en données temporelles revient au lecteur alors qu'avec la transcription chronologique, elle est assumée par le producteur ; dans le cas de la transcription linéarisée, elle est partagée. C'est la liberté laissée au lecteur qui semble expliquer le succès de la diplomatique au dépend des tentatives de chronologisation.

Les trois types présentés ci-dessus ne forment pas des catégories étanches. Un continuum existe, par exemple, entre la transcription diplomatique et la transcription linéarisée, laissant ouverte une gamme de formes intermédiaires, souvent adoptées. Ces transcriptions dites « semi-diplomatiques » – que

A. Grésillon (1994) appelle «mixtes», P.-M. de Biasi (2000) «semi-diplomatiques codées» et que P. D'Iorio (2010) nomme simplement «diplomatiques» réservant l'appellation «ultra-diplomatique» pour la reproduction à l'identique du manuscrit[4] – peuvent emprunter en proportions variables à l'un ou l'autre type de transcription. C'est le cas des conventions retenues pour le corpus d'écrit scolaire du projet *écriscol*.

Soit cet extrait d'une rédaction d'un élève de CM2[5] :


D'où vient-il? David le chef de la colo des Loutres de Honfleur.
appele Antoine et Sandrine colonie
Ils entres avec Antoine et Sandrine (deux
autres enfants) dans le jardin de la villa
inutile
bleu, tout doucement. Sandrine regarde dans la

Voici une transcription diplomatique de cet extrait (une différence de police et de graisse signale l'écriture de l'enseignant) :

D'où vient-il ? David le chef de la colo des Loutres de Honfleur
appele Antoine et Sandrine colonie
Ils entres ~~avec Antoine et Sandrine (deux~~

~~autres enfants~~) dans le jardin de la villa
inutile
bleu, tout doucement. Sandrine regarde dans la

Une transcription linéarisée du même passage :

<**D'où vient-il ?**> David le chef de la colo <colonie> des Loutres de Honfleur <appele Antoine et Sandrine > <Ils> entre<s> ~~avec Antoine et Sandrine (deux autres enfants)~~ <**Inutile>** dans le jardin de la villa bleu, tout doucement. Sandrine regarde dans la

[4] Ces variations en disent long sur le travail terminologique et réflexif que la génétique textuelle a encore devant elle.

[5] Cette image comme toutes celles que nous reproduirons par la suite sont issues du corpus *écriscol*.

Et un exemple de transcription chronologique simplifiée :

> [T1 – élève] : David le chef de la colo des Loutres de Honfleur entre avec Antoine et Sandrine deux autres enfants dans le jardin de la villa bleu, tout doucement. Sandrine regarde dans la
>
> **[T2 – enseignant]** :
> **D'où vient il ?**
> colo
> **(**deux autres enfants**)**
> **inutile**
> bleu
>
> [T3 – élève] : David le chef de la colonie des Loutres de Honfleur appele Antoine et Sandrine. Ils entres dans le jardin de la villa bleu, tout doucement. Sandrine regarde dans la

Voici, enfin, la transcription du même passage avec les conventions retenues par le projet *écriscol*, soit :
-<xxx> codant un ajout
-[xxx] codant une suppression
-σ# codant un soulignement
-P# codant l'écriture du professeur
-T2# codant un second moment d'écriture
-{xxx} codant un élément verbal ne s'inscrivant pas tel quel dans le texte.

> {P#D'où vient-il ?} {Pσ#David} le chef de la {Pσ#colo} <T2#colonie> des Loutres de Honfleur <T2#appel[e] Antoine et Sandrine> <T2#Ils> entre<T2#s> [T2#avec Antoine et Sandrine <P#(> {Pσ#deux autres
> enfants}] <P#)> {P#inutile} dans le jardin de la villa
> ble{Pσ#u}, tout doucement. Sandrine regarde dans la

Il s'agit, on le voit, d'une transcription linéarisée semi-diplomatique. Elle est en partie diplomatique parce qu'elle reproduit l'interruption des lignes liées au support et cherche à transcrire tous les signifiants graphiques du manuscrits ; elle est linéarisée car elle insère chaque évènement d'écriture dans une chaine de caractères. Les nécessités liées à la préparation du texte

pour la lemmatisation et l'annotation ont contraint à traiter certaines « ponctuations liées »[6] comme le soulignement ainsi que la rature par des signes discrets dans la ligne de caractères qui encadrent le segment textuel souligné ou barré. Si ces contraintes ne facilitent pas la lisibilité de la transcription, ce qui frappe surtout dans les choix effectués, c'est l'hétérogénéité des informations notées par les codages. En effet, les conventions retenues transcrivent non seulement des caractéristiques sémiotiques et topographiques du manuscrit ([xxx] pour un texte raturé, P pour le changement de graphie, <xxx> pour un texte en dehors de la ligne…) mais ils livrent aussi des indications chronologiques (T2) et discursives ({xxx} pour « texte commentatif »).

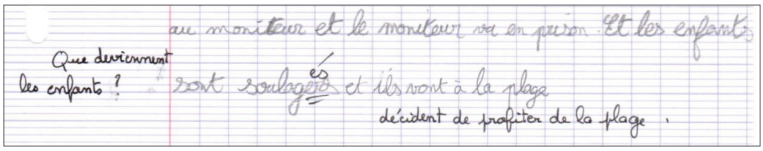
Ces deux éléments hétérogènes ont vraisemblablement été introduits pour affiner le codage sur deux aspects précis. Ainsi l'intégration d'une convention T2 (Temps 2) semble correspondre à la volonté de distinguer les opérations d'écriture faites au fil de la plume des opérations effectuées après une interruption. Un passage comme

sera transcrit « Il se releva au bout de quelque heure. Rogue [donna] <avait donné> » sans T2 pour coder le remplacement, car rien n'indique qu'il soit intervenu à un moment distinct de l'écriture du segment initial. Dans le passage précédent, en revanche, les remplacements ont été effectués par l'élève à cause des souligne-

[6] On entend par « ponctuation liée » les ponctuants non discriminés qui modifient les caractères d'écriture comme les majuscules, l'italique, le soulignement, le gras… La caractérisation de ces signes spécifiques pose problème. Certains linguistes ne les intègrent pas dans les signes de ponctuation à proprement parler : c'est le cas Catach (1994) qui les range parmi les « signes typographiques associés ». La plupart des spécialistes modernes néanmoins les intègrent pleinement dans leur description du système ponctuationnel : Tournier (1980) parle de « signes de ponctuation associés (ou liés) » en opposition aux « signes autonomes », Anis (1988) de « topogrammes liés » et de « topogrammes libres », Favriaud (2014) de « ponctuation grise » à côté d'une ponctuation blanche et noire. Dürrenmatt (2015), enfin, utilise plus volontiers le terme générique de « ponctuant » pour éviter la réduction de la ponctuation aux signes autonomes induite par le syntagme « signe de ponctuation ».

ments et des commentaires de l'enseignement sur le texte initial : ils prenaient par conséquent place à un moment d'écriture distinct. En somme, la convention « T2 » sert à établir la distinction, lorsqu'elle est possible, entre ce que les généticiens appellent les « variantes d'écriture » – les substitutions qui interviennent au moment de l'écriture – et les « variantes de lecture » – résultat des opérations effectuées par le scripteur après une pause, dans un moment de relecture du déjà écrit.

De même, les signes { } servent à discriminer les annotations métadiscursives de l'enseignement de ses ajouts destinés à être repris par l'élève dans une version ultérieure. Par exemple dans le segment suivant :



l'une des interventions sera transcrite {Que deviennent les enfants ?}, et l'autre <P#décident de profiter de la plage.>, c'est-à-dire comme un ajout.

La transcription adoptée pour le corpus *écriscol* emprunte, en définitive, aux trois grands types de transcription : elle fait cohabiter dans une transcription linéarisée des codages traditionnellement diplomatiques et d'autres d'ordre chronologique. Elle est exemplaire des paramètres multiples – contraintes technologiques, souci de la matérialité de l'écrit, recherche d'une analyse linguistique la plus fine possible – qui interviennent dans les choix de transcription et entrent en tension avec l'inscription dans un modèle unique.

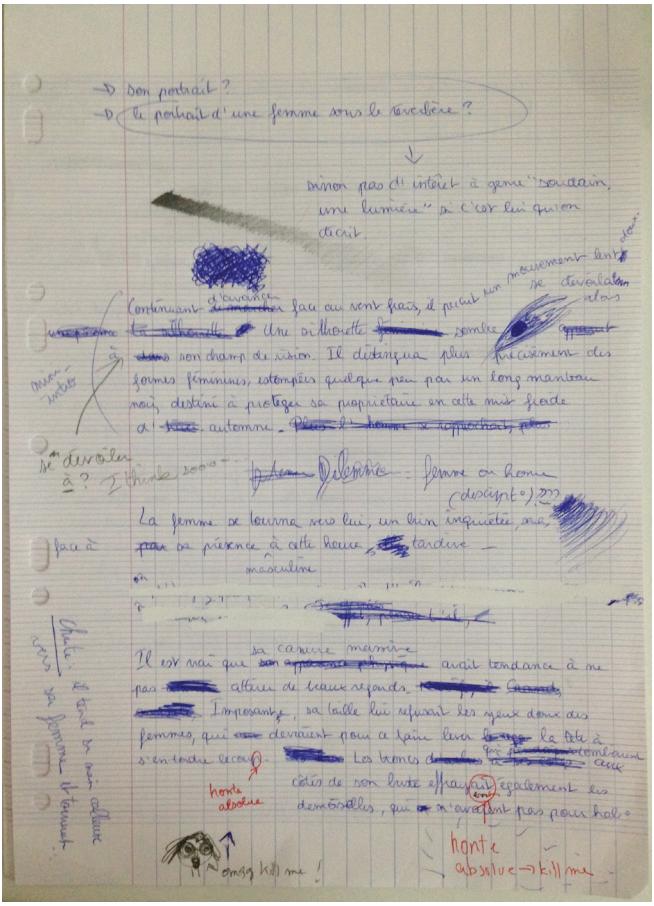
3. Quelques problèmes récurrents posés au transcripteur

Quels que soient les choix effectués, toute transcription est, comme le note A. Grésillon, à la fois plus riche et plus pauvre que le manuscrit dont elle est issue :

> Plus riche, car elle reflète un certain travail d'analyse : pour reproduire il ne suffit pas de copier, il faut d'abord *comprendre* les signes graphiques et les traduire en opération d'écriture. Plus pauvre, car toute transcription dactylographiée perd

> irrémédiablement ce que l'écriture manuscrite véhicule comme charge affective (hâte, blocage, angoisse, jubilation) et comme indice sur les mouvements scripturaux (changement de l'épaisseur du trait, alternances encre/crayon, noir/couleur, changement de ductus, etc.). (Grésillon 1994 : 129).

La différence sémiotique entre l'écriture manuscrite et l'écriture typographique n'est réductible qu'au prix de ces pertes. Ce sont les résistances que les écrits scolaires opposent au geste de transcription qui nous intéresseront à présent. On s'appuiera pour cela sur une page de brouillon d'un élève de seconde pour une écriture d'invention. Il s'agit d'un exercice de « continuation de texte ». Voici la page de brouillon :



Cette page dont la complexité, on le voit, n'a rien à envier aux manuscrits médiévaux ou aux manuscrits d'écrivains, servira à dégager certains problèmes récurrents posés aux transcripteurs.

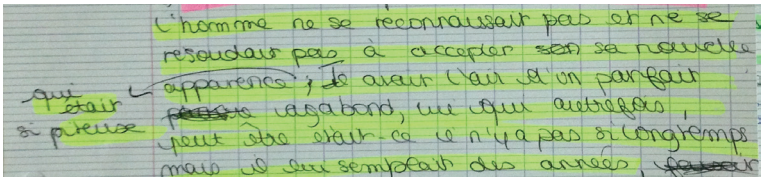
3.1 Le non verbal

La dimension iconique du manuscrit est un défi à la transcription. Dans les manuscrits d'écrivains comme, on le voit, dans ceux d'élèves cohabitent parfois plusieurs systèmes sémiotiques : signes verbaux, mais aussi symboles, codes mathématiques, tableaux, dessins, jeux de couleurs… Leur transcription pose problème, spécifiquement dans le cadre de la constitution d'un corpus informatisé. Les différentes solutions envisageables – reproductions informatiques, balises invitant à consulter l'original, intégration d'images dans la transcription, etc. – sont aussi couteuse en termes d'espace que de lisibilité. Pour autant, le non verbal ne saurait être évacué sans une perte d'information considérable car il renseigne sur le processus d'écriture et il participe directement de la signifiance textuelle. Comme l'écrit Louis Hay (1989 : 23), sur un manuscrit « tout est écriture, mais tout n'est pas verbe ».

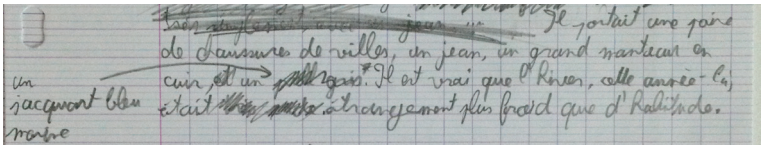
Les dessins, par exemple, ne sont pas toujours cantonnés en marge des brouillons comme des décharges nerveuses à la difficulté d'écrire. Ils peuvent aussi s'imbriquer étroitement à la construction du discours comme l'attestent les manuscrits de Valéry, de Faulkner, de Perec ou d'Hugo. C'est le cas également dans le brouillon ci-dessus où les ratures deviennent dessins et où la série de traits, au centre du manuscrits soulignent l'interrogation : « Dilemme : femme ou homme (descript°) ???? ». Ces éléments graphiques peuvent recouvrir une valeur modale comme les traits qui entourent l'inscription en rouge « honte absolue → kill me » au bas de la feuille pour souligner l'indignation malicieuse de l'élève face à ses propres erreurs orthographiques. Le tracé des lettres peut jouer la même fonction : par exemple celui, tremblant, du « Dilemme » au centre de la page. Autre élément sémiotique précieux à conserver : les différentes couleurs de tracé qui sont la trace de différents moments d'écriture. Ici, on distingue trois couleurs d'écriture : encre bleue, mine et encre rouge, les deux dernières correspondant à un moment de relecture et de correction.

Les éléments iconiques des manuscrits sont d'autant plus délicats à traiter que, contrairement aux signes verbaux, ils ne

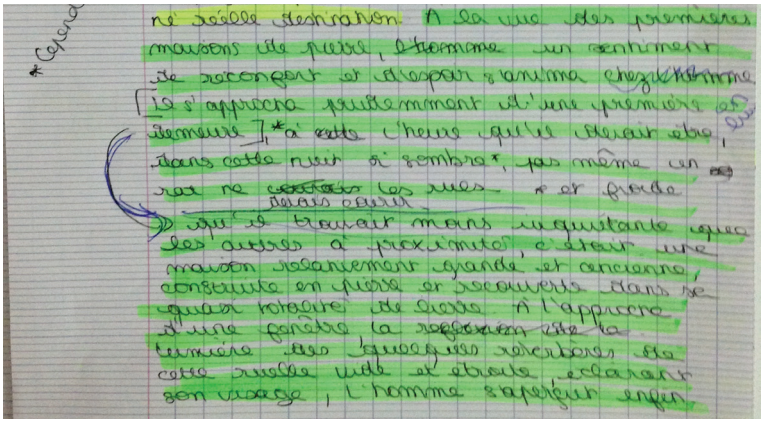
font tous pas nécessairement l'objet d'un transcodage univoque. Un même symbole peut être codé différemment selon sa fonction scripturale. Prenons l'exemple des flèches. Dans certains cas, les flèches sont un outil graphique pour une des opérations d'écriture, aussi bien l'ajout :



le remplacement :



que le déplacement :



Dans le cadre d'une transcription linéarisée ou semi-diplomatique, comme celle adoptée pour le corpus *écriscol*, ces flèches ne recevront pas de codage spécifique car les conventions de la suppression et de l'ajout suffisent à transcrire ces opérations. En vertu d'un principe d'économie, le remplacement, comme le déplacement dont il est une variante, sera rendu par la combinaison des conventions de la suppression et de l'ajout.

Les flèches présentes dans la page de brouillon présentée ci-dessus ont une autre fonction. Pour celles écrites en rouge ou à la mine, elles servent à relier les commentaires – « se dévoiler à ? » et les deux « honte absolue » – au segment textuel sur lequel ils portent. Ces flèches à valeur commentative ne seront pas transcrites telles quelles dans le corpus *écriscol* mais seront rendues, on l'a vu, par le symbole { }. Quant à la flèche en haut du manuscrit qui relie « le portrait d'une femme sous le réverbère ? » à « sinon pas d'intérêt à genre "soudain une lumière" si c'est lui qu'on décrit », elle recouvre encore une autre fonction. Elle ne marque pas une relation métatextuelle entre les deux phrases : elle est ici une notation abréviative de la transition entre deux idées. À ce titre, elle sera codée en tant que telle ou, à défaut, non marquée dans la transcription par la transition des deux phrases. Pour un même symbole, en somme, au moins trois transcriptions différentes.

3.2 Topographie de l'écrit

Même en se concentrant sur les signes verbaux de multiples problèmes surgissent, au premier rang desquels celui de leur disposition sur l'espace de la page. Les manuscrits modernes présentent en la matière une riche palette de possibilités : depuis le sage alignement des signes graphiques jusqu'à leur éclatement complet et désordonné sur l'espace de la page. L'écriture en milieu scolaire, prise dans un jeu de contraintes et d'évaluations et menée sur des supports le plus souvent lignés ou quadrillés, favorise la linéarisation de l'écrit. Un manuscrit à l'écriture linéaire comme c'est le cas de notre brouillon peut néanmoins poser des réelles difficultés de transcription, liées par exemple à l'orientation de l'écrit. Ainsi en va-t-il de l'inscription en bas à gauche de la page « Chute : il tend sa main calleuse vers sa femme et tournat » écrite perpendiculairement aux autres lignes d'écriture : où l'insérer dans le cadre d'une transcription linéarisée ? Il y aurait une part d'arbitraire à l'insérer à droite de tel ou tel segment comme il y aurait une lacune à ne pas signaler d'une façon ou d'une autre dans la transcription son orientation différente de toutes les autres traces écrites de la page.

On tâchera donc d'intégrer la topographie de l'écrit à la transcription, conventionnellement ou de façon mimétique, surtout si celle-ci relève de la structuration textuelle. Certains sauts

de lignes, modification d'interlignes, déplacement de marges ou alinéas correspondent, en effet, à des choix discursifs. On le remarque dans ce brouillon où la topographie de l'écrit dessine des régions discursives : le récit en construction est aligné à gauche sur la marge rouge alors que toutes les remarques métadiscursives sont soit écrites en marge soit centrées au centre de la page (par ex. « Dilemme »). La disposition de l'écrit sur l'axe horizontal de la page instaure ici des relations metatextuelles, tandis que la disposition sur l'axe vertical correspond à une organisation de l'ordre syntagmatique du texte. Si l'on confronte le brouillon à sa mise au net correspondante par le même élève, il apparaît que les sauts de lignes et la création d'ilots textuels dans le brouillon dont certains semblent naître par accident – causé le plus souvent par des ratures – ont une incidence directe sur la mise en paragraphe dans la version ultérieure. La gestion de l'espace graphique est aussi gestion du discours.

3.3 Ratures et autres suppressions

S'il est un objet que la génétique textuelle a particulièrement étudié d'un point de vue à la fois empirique et théorique, c'est bien la « rature ». « L'univers de la rature » (Biasi 2000 : 53) a fait l'objet de très nombreuses études qui ont exploré sa fonction métalinguistique ou encore ses variétés formelles et fonctionnelles. La rature ne saurait pourtant résumer à elle seule l'opération de suppression. Il existe d'autres gestes, alternatifs à l'acte de barrer, à même de retrancher du déjà écrit : découper un passage, le blanchir, gommer, coller un papillon… Mais, comme le note Claire Bustaret, ces variantes sémiotiques ne sont pas intégrées à la théorie génétique :

> Les opérations d'écriture définies par la critique génétique, ajout, suppression et déplacement, ne comportent pas de spécification matérielle : insérer une phrase en l'inscrivant en marge ou en interligne, ou bien en procédant par collage, ne saurait pourtant être décrit comme strictement équivalent, ni du point de vue de l'action, ni du point de vue de résultat – et il en va de même pour le découpage, qui agit autrement que la rature. (Bustarret 2011 : 354)

L'importance de ces différentes caractéristiques matérielles de la suppression apparaît dans le cas de notre brouillon. Au centre de la page, figurent deux lignes aux trois quarts blanchies : comment les transcrire ? On rendra vraisemblablement ce segment par une convention indiquant un passage barré et illisible. Or, que faire du passage barré et lisible à la fin de la seconde ligne qui n'a pas été blanchi ? Et comment discriminer un passage supprimé rendu illisible par l'usage du blanc ou par l'acte de barrer ? Il paraît évident que ni ces deux gestes ni les traces qui en résultent ne sont équivalentes pour le sujet écrivant : l'usage unique du blanc sur cette page de brouillon par ailleurs bien raturé devrait donc être restitué dans la transcription. Si la rature en ce qu'elle est de l'annulé lisible se prête à la transcription, il convient aussi de s'interroger sur les manières de restituer ces segments annulés, non nécessairement lisibles, dans leurs différences matérielles.

3.4 Le métadiscursif

Lorsque l'on transcrit, la tentation est grande de postuler ce qui serait « le texte » – ici, le récit qu'élabore l'élève à partir de la contrainte – et ce qui serait du « hors-texte », auquel on pourrait donner des noms variés : gloses, commentaires, *marginalia*… Cette division instinctive correspond au constat que sur le manuscrit cohabitent des données hétérogènes, qu'on peut ranger dans deux catégories : des données visant directement la production d'un texte linéaire, destinées à être repris dans une version ultérieure, et des éléments variés (notes de régie, commentaires, liste de mots…) qui accompagnent cette production. Comme la frontière entre ces deux ordres de données n'est pas toujours nette, un principe d'exhaustivité préside : tout ce qui est écrit sur le manuscrit doit être transcrit car tout sur le manuscrit fait texte.

Même si elles font obstacle à la linéarisation, les annotations métadiscursives doivent d'autant plus être conservées qu'elles sont le témoignage de la réflexivité énonciative du scripteur. La richesse de ces annotations dans le cas de notre brouillon est évidente. On y trouve en effet, des interrogations sur la langue – « se dévoiler à ? » – et des retours sur le déjà écrit avec des commentaires sur les fautes d'orthographes. D'autres

interrogations portent sur la construction du texte, soit en amont de la production :

→ son portrait ?
→ le portrait d'une femme sous le réverbère ?
↓
sinon pas d'intérêt à genre "soudain une lumière" si c'est lui qu'on décrit

soit en interrompant le fil de la production :

~~Dilemme~~ Dilemme : femme ou homme (descript°) ????

Ces annotations relèvent des notes de régies c'est-à-dire des instructions métascripturales que le scripteur s'adresse à lui-même pour la production de son texte. Elles sont ici sous formes nominales et ont la particularité de recourir au vocabulaire technique : « mini-intro », « description », « chute »… Un autre aspect intéressant est le changement de langue qui s'y fait jour. C'est lors du retour sur le déjà écrit que l'élève passe à l'anglais : « I think so », écrit-il, lorsqu'il s'interroge sur la construction syntaxique du verbe dévoiler ; « kill me » lorsqu'il repère des fautes orthographiques. Le changement linguistique est aussi un changement de registre ; il indique un rapport distancié face à sa propre production lié manifestement à un temps différent, celui de la révision. Les opérations métadiscursives, et les annotations qui les accompagnent, sont constitutives de la production de tout écrit un peu élaboré. Leur conservation dans la transcription est donc essentielle pour l'analyse des processus de production.

Conclusions

La richesse des annotations discursives présentes sur cette page de brouillon attestent que loin d'être simples les écrits scolaires soulèvent des problèmes similaires à ceux posés par certains manuscrits d'écrivains. Claire Doquet (2013) dans un article qui fait le point sur l'approche génétique appliquée aux écrits d'élèves, met en garde contre ce qui serait un applicationnisme étroit. Il s'agit, en effet, ni de plaquer sur les écrits d'élèves les catégories dégagées de l'analyse des corpus littéraires, ni de rechercher dans les écrits scolaires des phénomènes identiques, au même ni-

veau d'analyse, que les manuscrits d'écrivains. S'ils témoignent de degrés d'expertise incomparables, le manuscrit d'un écrivain et celui d'un élève ont ceci de commun qu'ils gardent la trace d'une mise en jeu d'un rapport à la langue et à l'écriture. Les analyses de ce rapport sont certes conditionnées par la transcription mais ils sont déjà à l'œuvre dans cette opération. C'est pourquoi l'écrit d'un d'élève appelle la même rigueur que tout autre manuscrit moderne. Une transcription consciente de ses enjeux, qui explicite ses choix en fonction de ses objectifs et de son cadre théorique, est le préalable à toute analyse.

Bibliographie

Anis J. (1983). « Vilisibilité du texte poétique », *Langue française* 59 : 88-102.

Anis J. (1988). *L'écriture. Théories et descriptions*, avec la collaboration de J.-L. Chiss et C. Puech. Bruxelles : De Boeck.

Arabyan M. (2012). *Des lettres de l'alphabet à l'image du texte. Recherches sur l'énonciation écrite*. Limoges : Lambert-Lucas.

Bellemin-Noël J. (1972). *Le texte et l'avant-texte : les brouillons d'un poème de Milosz*. Paris : Larousse.

Biasi de P.M. (2000). *La génétique des textes*. Paris : Nathan.

Biasi de P.M. (1996). « Qu'est-ce qu'une rature ? », in B. Rougé (éd.) *Ratures et repentirs*. Pau : Publications de l'Université de Pau, 17-47.

Biasi de P.M. et Herschberg Pierrot A. (éd.) (2015). Littérature 178 « Génétique : les chemins de la création ».

Bustarret C. (2011). « Couper, coller dans les manuscrits de travail du XVIIIe au XXe siècle », in C. Jacob (éd.) *Lieux de savoir 2. Les mains de l'intellect*. Paris : Albin Michel, 353-375.

Catach N. (1994). *La ponctuation*. Paris : Presses Universitaires de France.

Cerquiglini B. (1989). *Éloge de la variante. Histoire critique de la philologie*. Paris : Seuil.

Cerquiglini B. (1998). « Variantes d'auteur et variance de copiste », in L. Hay (éd.) *La naissance du texte*. Paris : José Corti, 105-120.

Cerquiglini B. (2010). « Vingt ans après ». *Genesis* 30 : 15-17.

Crasson A. et Fekete J.-D. (2004). « Structuration du manuscrit : du corpus à la région », *Proceedings of CIFED* 2004 : 162-168 [En ligne : http://www.item.ens.fr/index.php?id=173027].

Crasson A. et Hay L. (éd.) (2013). *Genesis* 37 : « Verbal – non verbal ».

Doquet C. (2013). « Ancrages théoriques de l'analyse génétique des textes d'élèves », in C. Boré et E. Calil (éd.) *L'école, l'écriture et la création. Études françaises et brésiliennes*. Louvain-la-Neuve : Academia-L'Harmattan, 33-53.

Dürrenmatt J. (2015). *La ponctuation en français*. Paris : Orphys.

Duval F. (2007). « À quoi sert encore la philologie ? », *Laboratoire italien* 7 : 17-40 [En ligne : http://laboratoireitalien.revues.org/128].

Espagne M. (2010). « Philologie et critique génétique », *Genesis* 30 : 19-20.

Favriaud M. (2014). *Le plurisystème ponctuationnel français à l'épreuve de la poésie contemporaine*. Limoges : Lambert-Lucas.

D'Iorio P. (2010). « Qu'est-ce qu'une édition génétique numérique ? », *Genesis* 30 : 49-53.

Fuchs C. *et alii* (1982). *La genèse du texte : les modèles linguistiques*. Paris : Édition du CNRS.

Grésillon A. (1994). *Éléments de critique génétique*. Paris, Presses Universitaires de France.

Grésillon A. (1996). « Raturer, rater, rayer, éradiquer, radier, irradier », *in* B. Rougé (éd.) *Ratures et repentirs*. Pau : Publications de l'Université de Pau, 49-60.

Hay L. (1985). « "Le texte n'existe pas" : réflexions sur la critique génétique », *Poétique* 62 : 146-158.

Hay L. (1989). « L'écrit et l'imprimé », *in* L. Hay (éd.) *De la lettre au livre. Sémiotique des manuscrits littéraires*. Paris : Éditions du CNRS : 7-34.

Hay L. (éd.) (1996). *Genesis* 10 « Sémiotique ».

Lebrave J.-L. (1984). « Le traitement automatique des brouillons », *Programmation et Sciences de l'homme*, LISH, CNRS, n° 3.

Lebrave J.-L. (1986). « L'écriture interrompue : quelques problèmes théoriques », in L. Hay (éd.) *Le manuscrit inachevé. Écriture, création, communication*. Paris : Éditions du CNRS : 127-165.

Lebrave J.-L. (1990). « Déchiffrer, transcrire, éditer la genèse », in A. Grésillon, J.-L. Lebrave et C. Viollet (éd.) *Proust à la lettre : les intermittences de l'écriture*. Tusson : Du Lérot, 141-205.

Lebrave J.-L. (1992). « La critique génétique : une discipline nouvelle ou un avatar moderne de la philologie ? », *Genesis* 1 : 33-72.

Lebrave J.-L. (2009). « Manuscrits de travail et linguistique de la production écrite », *Modèles linguistiques*, 59 : 13-21.

Masai F. (1950). « Principes et conventions de l'édition diplomatique », *Scriptorum* 4/2 : 177-193.

Rey-Debove J. (1982). « Pour une lecture de la rature », in C. Fuchs *et alii* (éd.) *La genèse du texte : les modèles linguistiques*. Paris : Éditions du CNRS, 103-127.

Segre C. (1995). « Critique des variantes et critique génétique », *Genesis* 7 : 29-45.

Segre C. (2010). « Philologie italienne et critique génétique », *Genesis* 30 : 25-27.

Tournier C. (1980). « Histoire des idées sur la ponctuation », *Langue française* 45 : 28-40.