HAL open science

LA DYNAMIQUE D'ÉCRITURE DANS LA DESCRIPTION LINGUISTIQUE. NOUVEAUX MODES DE VISUALISATION DE L'ÉCRITURE ENREGISTREE

Christophe Leblay, Gilles Caporossi

To cite this version:

Christophe Leblay, Gilles Caporossi. LA DYNAMIQUE D'ÉCRITURE DANS LA DESCRIPTION LINGUISTIQUE. NOUVEAUX MODES DE VISUALISATION DE L'ÉCRITURE ENREGISTREE. Dossiers d'HEL, 2016, Écriture(s) et représentations du langage et des langues, 9, pp.180-193. ⟨hal-01304945⟩

HAL Id: hal-01304945

https://hal.science/hal-01304945v1

Submitted on 20 Apr 2016

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.

Autorisation HAL

HAL Authorization

LA DYNAMIQUE D'ÉCRITURE DANS LA DESCRIPTION LINGUISTIQUE. NOUVEAUX MODES DE VISUALISATION DE L'ÉCRITURE ENREGISTREE

Leblay Christophe,
Université de Turku & TERs, ITEM, CNRS-ENS
Caporossi Gilles,
HEC Montréal & TERs, ITEM, CNRS-ENS

RÉSUMÉ
Ce travail est centré sur l'écriture, comme processus linguistique descriptible et reproductible, s'inscrivant dans le temps et l'espace, c'est-à-dire dans l'avant d'un texte finalement stabilisé. En adoptant une approche génétique du texte, associée à un programme d'enregistrement de l'écriture, est décrit et exemplifié un nouveau mode de visualisation basé sur une transposition de la théorie mathématique des graphes. Cette visualisation, bâtie à partir des documents sources que sont les enregistrements (logs) exhaustifs des activités réalisées à l'aide du clavier et des mouvements du curseur, fait apparaître nettement des phases temporelles dans l'écriture.
En d'autres termes, il s'agit de s'interroger pour savoir comment la description linguistique contemporaine peut tirer profit de ces nouveaux modes de visualisation.

MOTS-CLEFS
Description linguistique – génétique de l'écriture – graphe mathématique – temporalité de l'écriture – visualisation

ABSTRACT
This article focuses on writing as a process. The main purpose of this work is to show what kind of phenomena appear in space and time before the final text is stabilized. By adopting the genetic criticism approach associated with a keystroke logging program, is described and exemplified a new display mode based on a transposition of the mathematical theory of graphs. This visualization, built from source documents (logs) that are comprehensive activities using the keyboard and cursor movements, clearly shows the temporal phases in writing. The work shows that using graphs as a mode of visualization of writing process can be quite promising. Contemporary linguistic description can take advantage of these new modes of visualization.

KEY-WORDS
Genetic criticism – linguistic description – mathematical graph – writing temporality – visualisation

Dynamique et visualisation de l'écriture enregistrée

Préférer une approche centrée sur la production, plutôt que sur le produit, implique un choix technologique très particulier propre à rendre compte du dynamisme de cette production ; autant le choix du papier que celui de l'écran sont des choix relevant d'une technologie, ces deux choix étant valides pour mettre en relief une dynamique de l'écriture. Tous deux révèlent, à leurs manières, certaines caractéristiques de l'écriture qui ne s'excluent pas les unes les autres. L'idée sous-jacente est alors que chaque apport technologique permet de montrer un aspect de l'écriture, l'ensemble, encore bien ouvert, des technologies permettant d'offrir un panorama de plus en plus complet des phénomènes scripturaux. Depuis le tout premier numéro (Grésillon et Lebrave 1983), et plus précisément depuis maintenant dix ans, les revues *Langages (No 146, 164 et 177)*, *Langue française* (No 155) et *Modèles linguistiques* (Tome XXX, 59) ont entrepris une large diffusion des recherches consacrées aux différentes approches de la production écrites. Nous nous inscrivons dans cette continuité.

Il n'est guère possible de se satisfaire de termes conçus pour la description de phénomènes linguistiques de réception à commencer par la notion de *texte*, corrélative à celle de produit. Celle-ci semble, comme le souligna, déjà en 1982, Bellemin-Noël (p. 161), être faite davantage pour des approches en réception, en lecture donc, que pour des approches en production, en écriture : il s'agit, dans les approches de réception, de donner du sens à ce qui est (déjà) écrit indépendamment des conditions d'énonciation. Il est alors inévitable de s'interroger sur la pertinence descriptive de notions qui, comme celle du texte, n'ont pas été forgées pour l'étude de phénomènes de production écrite. Comme l'exprime judicieusement Ferrer (2008, p. 190), il nous faut apprendre à faire face à des phénomènes qui

> ne [sont] pas des textes [...]. Ça ressemble à du texte, mais ce n'est pas du texte. C'est quelque chose qui sert à <u>fabriquer</u> du texte. (Nous soulignons).

Notre approche présuppose donc un cadre théorique et méthodologique, selon lequel le produit n'est pas l'image fidèle de la production : chercher à comprendre les processus d'écriture-réécriture (*planification*, *textualisation* et *révision)* à partir du seul texte produit serait une pure illusion.

Il est nécessaire d'insister sur l'idée que la prise en compte de la relation chronologique est la condition *sine qua non* pour redonner aux études de production écrite leur véritable statut d'*études de production*. A l'instar des approches en production orale, il est essentiel de pouvoir être capable d'appréhender l'écriture dans son déroulement temporel, et pour ce faire, d'enregistrer l'écriture. Pour reprendre une expression de Claudine Fabre-Cols, pionnière dans la description de la production écrite, l'écriture est « *fondamentalement dans le temps* » (2004, p. 21). Mais c'est à Aragon (1979, p. 9) que nous empruntons l'expression de *temps de l'écriture*, lequel déclara, évoquant les dons qu'il avait fait de ses manuscrits aux chercheurs, dans son allocution au CNRS :

> « Ne fallait-il pas mettre à la disposition de ceux qu'on appelle les chercheurs non seulement <u>l'écrit figé</u> par la publication, mais le <u>texte en devenir</u>, saisi pendant <u>le temps de l'écriture</u>, avec ses ratures comme ses repentirs, miroir des hésitations de l'écrivain comme des manières de rêverie que révèlent les achoppements du texte ». (Nous soulignons).

Nous allons commenter ce qu'implique, en termes de méthodologie, le passage d'une technologique à l'autre (1), afin de pouvoir ensuite mettre en évidence l'importance que représente la possibilité d'enregistrer la dynamique de l'écriture (2) ; ces précautions achevées, nous pourrons alors développer ce que la théorie mathématique des graphes peut apporter dans la recherche d'une visualisation de la dynamique d'une écriture préalablement enregistrée (3).

Christophe Leblay, Gilles Caporossi

1. Du papier a l'écran- d'une technologie à l'autre

Il ne s'agit pas d'opposer un support à un autre, le papier à l'écran, en prétendant que le second représenterait une avancée technologique. Les deux sont des supports de haute technologie, à n'en pas douter. Il est donc préférable de parler de deux technologies, de celle du papier, comme de celle de l'écran. Regardons d'abord certaines caractéristiques du support papier, avant de considérer le support numérique.

1.1 Brouillons sur papier

Voici, pour illustrer notre propos, un fac-similé de la première page du manuscrit « Sémiologie de la langue » de l'article (inclus dans les *Problèmes de Linguistique Générale*) d'Émile Benveniste. Grâce à ce premier document, il nous est possible de nous interroger sur la différence entre *manuscrit* et *brouillon*.

Dès la fin des années 70, Bellemin-Noël, lors du colloque international de *textologie,* tenu à Mátrafüred (Hongrie), s'interrogeait publiquement sur une difficulté terminologique de taille : deux mots étaient en concurrence pour désigner les documents de rédaction disponibles, le premier étant *manuscrit*, le second, *brouillon*. Pour Bellemin-Noël (1982, p. 161-162), ces deux termes se différenciaient de la manière suivante : le manuscrit désignait, « *en toute rigueur, des feuilles de papier couvertes de signes tracés par la main de l'écrivain* » ; il pouvait sans difficulté permettre « *d'évoquer le support tangible de la rédaction, voué à la conservation par la complaisance de l'auteur et par l'admiration des amateurs de belles-lettres*. » A la différence du précédent, le terme de brouillon servait à désigner « *la matérialisation d'un discours inachevé, prospectif, parfois mis au rebut, le plus souvent transformé au fil d'une pratique d'élaboration* », en permettant de distinguer le pluriel, les brouillons, conçus comme ensemble de « *tout ce qui a servi à la composition d'un ouvrage, [de] ce qui n'a jamais eu le statut du publiable* » du singulier, le brouillon, compris comme « *le premier jet et ses métamorphoses (ajouts, corrections, ratures et substitutions) jusqu'à l'état final de la première publication* », et les transformations apportées par « *les modifications de l'imprimé* », nommées aussi *variantes*. Néanmoins, il est aisé de comprendre la gêne de Bellemin-Noël, pour un terme qui « *connote quelque chose d'embrouillé, ou de non débrouillé, des tâtonnements* », pour un terme qui « *implique que l'auteur marche vers une perfection dont il aurait la prescience, alors que pour lui, en réalité, les mots qu'il écrit sont d'emblée une formulation virtuellement acceptable ; c'est après coup qu'il se découvrira insatisfait, et reviendra sur son travail*. ». Pour toutes ces raisons, Bellemin-Noël (1982, p. 162) proposa un terme neutre, celui d'*avant-texte*[1]. Dans ce travail, nous reprenons le terme d'avant-texte, bien plus apte aux études de production. Nous en proposons un exemple, reproduit ci-dessous :

[1] Il n'est donc pas surprenant de voir que, presque 25 ans après, un numéro de la revue *Langue française* (155, «Avant le texte : les traces de l'élaboration textuelle ») publié en 2007 et coordonné par Lucille Chanquoy et Irène Fenoglio, reprend, dans son titre, la notion de l'avant du texte. Dans leur introduction, les auteurs reviennent explicitement sur la définition, en proposant comme définition : « *Tout ce qui se passe avant la production du texte écrit et dont les traces attestent d'une mise en acte cognitive et graphique directement liée au texte final produit*» (p. 5).

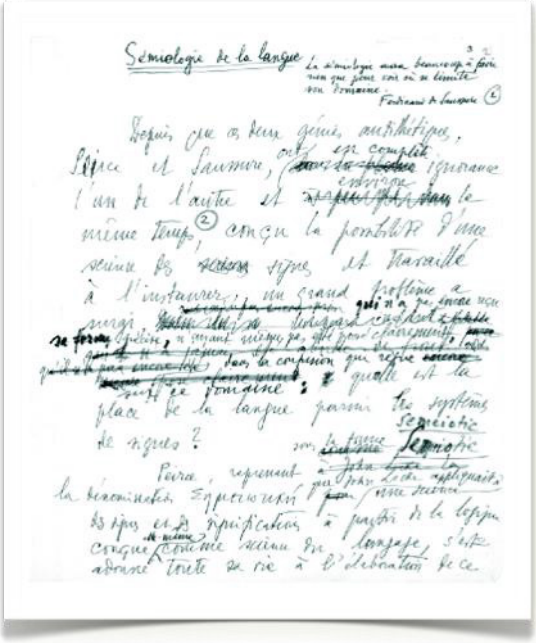

Figure 1. Première page de « Sémiologie de la langue »

Que montre, ou pas, un tel document ?
Tout d'abord, autant le manuscrit que le brouillon coexiste sur cette même page : il s'agit bien d'une *feuille de papier couverte de signes tracés* par la main de Benveniste tout autant qu'un *premier jet et ses métamorphoses (ajouts, corrections, ratures et substitutions)*. De ces deux statuts, le second est celui qui intéresse la description linguistique, le premier intéressant principalement la codicologie.
Ce manuscrit, devenu donc brouillon sous l'œil averti du linguiste, nous montre l'importance des ratures : il ne viendrait à personne l'idée de mettre en doute les compétences scripturales que peut avoir un auteur comme Benveniste. Il nous est donc facile d'apprendre des manuscrits de linguistes (Fenoglio 2012).
Néanmoins, ce brouillon ne peut pas nous montrer la chronologie des retours. Il peut y avoir des hypothèses, certes, mais rien que des hypothèses. Combien de temps Benveniste a-t-il pris pour écrire et réécrire ces lignes ? Dans quel ordre a-t-il procédé ?

1.2 Brouillons numériques
A la différence du précédent, le brouillon numérique est réalisé à l'aide d'un écran et le plus souvent d'un clavier[2]. Ces outils de production sont, eux aussi, des outils d'aide à l'écriture, puisqu'ils sont le plus souvent associés à l'utilisation d'un logiciel de traitement de texte. Si, néanmoins, ces logiciels sont très performants, ils ne sont pas ceux qui nous intéressent dans cette recherche qui prend en compte d'autres logiciels. Il est néanmoins clair que l'utilisation de l'écran ne peut rendre compte, à lui seul, du travail effectué pendant l'écriture. Voici, pour illustrer notre propos, un document numérique.

[2] Il arrive que l'écriture se fasse aussi à l'aide d'un stylo optique, ou stylet, sur une tablette graphique.

Environnement idéal

Qu'est-ce qui me vient en tête spontanément quand on me donne la tâche de décrire un environnement idéal? Beaucoup de verdure, peu de pollution, suffisamment d'espace pour que tout un chacun se trouve à l'aise avec soi-même et avec les autres. Le climat? Je dois dire que les saisons distinctes et variées typiques du Nord me plaisent, etv que c'est quelque chose que je voudrais avoir dans un environnement idéal. Autrement dit, un été chaud et ensoileillé sans oublier évidemment les journées lumineuses et extrêmement longues comme en Finlande en été, un hiver blanc, avec du soleil et de la neige, et encore deux saisons de "transmission", de passage on peut dire, pour passer du froid au chaud, ou inversement: le printemps et l'automne.

S'il n'est pas possible, pour moi, d'imaginer une vie idéale sans la nature telle que je la connais dans mon pays natale, il m'est tout aussi impossible d'imaginer une existence heureuse sans la présence des enfants et des animaux. Les enfants et les animaux font partie de la nature.
J'ai toutjours vécu au bord de la mer, ou plus précisément, je vis mes hivers au bord de l amer proprement dite et mes étés à la campagne au bord d'un grand lac. Cette présence de l'eau dans mon environnement est quelque chose qui me manque lorsque je suis

Figure 2. Page imprimée d'une écriture numérique

Que montre, ou pas, un tel document ?
Tout d'abord, ni le manuscrit ni le brouillon ne coexistent sur cette même page : les deux statuts semblent être neutralisés par un effet de mise au propre immédiat, représentatif d'un tapuscrit. À proprement parler, il n'y a plus de *feuille de papier couverte de signes tracés* par la main de son auteur, et pas davantage un *premier jet et ses métamorphoses (ajouts, corrections, ratures et substitutions)*. De ces deux statuts, le présent document ne semble en retenir donc aucun, contrairement au document précédent qui, lui, offrait, en couches superposées, les deux statuts.
Cette différence majeure, dans la représentation du travail en production écrite, nous amène à nous interroger sur la légitimité de telle ou telle représentation : ne serait-il pas alors plus pertinent de travailler à partir d'un support qui permette de voir de suite les traces laissées, comme le font le papier et le crayon ? Ce serait oublier deux arguments. Le premier est que le support papier ne donne que des traces de modifications spatiales, sans aucune possibilités de repérage temporel de ces traces. Le second argument concerne l'évolution des logiciels d'aide à l'écriture, tels le traitement de texte : le support numérique propose de plus en plus de solutions d'accès à ce qu'a été le déroulement temporel d'une écriture. Des fonctions, telles *Time Machine*, permettent de pouvoir revenir, à chaque instant, à une chaîne de versions antérieures. D'une manière générale, les traitements de texte ont intégré, systématiquement, la fonction de *révision* (écriture-réécriture). Cette évolution dans la manière de concevoir l'aide à l'écriture-réécriture reste déterminante pour l'avenir en plaçant au premier plan l'idée qu'il ne peut y avoir d'amélioration sans retour dans son écriture. Cette page, écrite sur un écran, donne donc accès au temps de l'écriture, tout simplement parce qu'il existe un outillage pour prendre en compte la succession des versions produites au cours du processus d'écriture, ce que ne peut pas faire une page écrite sur un papier.

1.3. Représentation versus visualisation
Arrivés au terme de ces remarques concernant les différents supports de brouillon et les implications qu'ils peuvent avoir dans la prise en compte des dynamiques de production, il nous

semble nécessaire d'attirer l'attention sur un terme, trop communément utilisé, celui de *représentation*. Sans aucun doute, un tel terme pourrait faire l'objet d'une réflexion séparée sur la représentation dans l'acquisition de l'écriture.
Nous employons ce terme dans le sens donné par la pragmatique : *représentation supposée partagée*. Il est simplement question d'un savoir commun minimum que doivent partager les interlocuteurs afin que puisse fonctionner l'intercompréhension. Le terme de *représentation* est alors très proche de celui de *schématisation*, développé par les travaux de Grize, où il s'agit de « *faire voir quelque chose à quelqu'un ; plus précisément, c'est une représentation discursive orientée vers un destinataire de ce que l'auteur conçoit ou imagine d'une certaine réalité* » (1996, p. 50). Or, c'est bien cette conception ou cette imagination de la réalité de l'écriture qui est loin d'être partagée. Nous sommes simplement habitués à voir de l'écriture sur une page raturée, pas ailleurs, pas autrement. Il reste donc un travail important à faire pour rendre légitimes d'autres représentations.
Les premiers travaux sur les représentations faites du processus d'écriture datent des années 1990, avec l'arrivée de programmes tels que *Genèse du texte* ; ces programmes seront à l'origine de travaux spécialisés dans la recherche sur la production écrite enregistrée, dite *en temps réel*. Parmi ces travaux, certains auront pour ambition de trouver des pistes dans la *notation* du processus d'écriture (Kollberg 1998), ancêtre des recherches sur la représentation numérique.

2. Représentations numériques standardisées

Les deux types de représentations suivantes sont toutes les deux issues d'un fichier d'enregistrement. Ces deux fichiers ont tendance à définir, depuis quelques années maintenant, des formats standards dans l'édition numérique des écritures enregistrées.

2.1 Le fichier d'enregistrement

Dans l'exemple qui suit, un début de production enregistré a donné ce qui suit :

time	type	from	to	key
0.00	10	1	0	<START>
4.21	7	0	0	L
4.46	7	1	1	I
4.75	7	2	2	E
5.05	7	3	3	U
5.26	7	4	4	
5.70	7	5	5	I
5.86	7	6	6	D
8.08	7	7	7	…
8.36	7	8	8	A
8.53	7	9	9	L
8.81	7	10	10	
12.28	5	11	11	<DELETE>
12.45	5	10	10	<DELETE>
12.61	5	9	9	<DELETE>
12.78	5	8	8	<DELETE>

Figure 3. Fichier d'enregistrement (Log)

Il convient de distinguer, parmi tous les termes qui se réfèrent, d'une manière ou d'une autre, aux opérations d'écriture (variante, inscription) celui d'*événement d'écriture*. Selon Strömqvist (2006, p. 46), l'événement est représenté par une série de cinq arguments (temps, type d'événement, position à l'écran, alternative entre touche ou souris). Pour simplifier à l'extrême, il est possible de dire qu'un événement représente une ligne du fichier d'enregistrement.
Mais, ce serait oublier qu'en se (re-)présentant ainsi, les *fichiers d'enregistrement* permettent de remodaliser cette troisième dimension dont parlait Grésillon (1994), la dimension du temps, cette

fois-ci saisie sur un support numérique : tout d'abord, ces fichiers d'un nouveau genre permettent un réexamen sans complaisance de ce que sont l'espace et le temps dans le processus d'écriture en permettant l'enregistrement séquentiel (log), dans un fichier dédié, de tous les événements affectant le processus d'écriture, depuis les frappes au clavier jusqu'aux manipulations de périphériques (souris et pavé tactile). Ensuite, ces événements sont datés et classés selon une chronologie absolue; ce qui permet d'analyser pas à pas l'activité du processus. Le log est bien cette suite d'événements organisés selon une chronologie absolue et formant un tout. Nous proposons de faire une différence notable entre *chronologie relative* et *chronologie absolue* : la seconde note une chronologie qui part du début de l'enregistrement, la première note le temps entre les différents événements. Cette notation chronologique oblige conséquemment à une *lecture séquentielle* du log, c'est-à-dire à une lecture, où chaque caractère suit chaque caractère, où chaque mot suit chaque mot, du début à la fin, dans un ordre temporel immuable (Fenoglio et Ganascia 2008).
Ces fichiers d'enregistrement restent donc le document numérique *princeps*, à partir duquel tout se fait. Ceux-ci peuvent être lus et vus 1) verticalement, sous la forme de colonnes de chiffres et de lettres, ou bien 2) horizontalement, sous la succession de lignes, chaque ligne représentant un événement. Il est déjà clair que le fichier d'enregistrement s'intéresse de près à la problématique de la *linéarisation* (Severinson-Eklundh 1994).

2.2 Le texte linéaire

Au sein de tout logiciel d'enregistrement (*Eye and Pen*, *Genèse du texte*, *Inputlog*, *Scriptlog*, *Translog*…), il est donc devenu maintenant commun d'associer au *fichier d'enregistrement* (*log*), un autre fichier, nommé lui, le plus souvent, *texte linéaire* (*linear text*). Voici, maintenant donc, la même production que celle montrée précédemment, mais sous le format d'un texte linéarisé.

<START>LIEU IDÉAL <DELETE11>Lieu idéal<CR><CR>J'aimerais vivre ci<DELETE>omme la plupart des Finlandais,<LEFT33>,<RIGHT33> à la campagne près d'un lac mais t´<DELETE2>pas très loin da la

Figure 4. Texte linéaire

Comme nous pouvons le voir, la représentation suivante, issue du fichier d'enregistrement (*Scriptlog, version Mac*) est une tentative de linéarisation de toute l'information donnée par le *log*. Cette représentation ne fait que linéariser ce que dit le *log*, d'une autre manière. Si elle reste plus satisfaisante pour l'œil volontaire du chercheur, elle l'est bien moins pour les autres.

3. La genèse représentée par un graphe

La théorie mathématique des graphes, approche qui date du XVIIIe siècle, a déjà montré sa capacité à s'adapter à des disciplines très diverses, comme par exemple, à la chimie, aux télécommunications, aux transports et aux réseaux sociaux.
Pourquoi avoir recours à cette théorie ? Principalement parce qu'elle permet une très grande souplesse dans l'encodage des données, simplement avec deux formes géométriquement simples, le cercle et le trait. Il s'agit alors de donner des valeurs à ces images que sont le cercle et le trait, ainsi que de s'interroger sur le fait de joindre, ou pas, un cercle à un autre cercle par l'intermédiaire de traits.
Ce travail de description de la dynamique de l'écriture à partir de graphes a déjà été initialisé (Caporossi et Leblay 2011). Nous présentons, dans les lignes qui suivent, à partir de saisies d'écran, des exemples de *patterns temporelles* d'opérations d'écriture. Mais regardons d'abord comment fonctionne un graphe, depuis sa construction jusqu'aux liens qui les composent.

3.1 Construction en *cellules*

Soit une très brève production où, dans un premier temps, est écrit, d'un seul jet, *Un chat mange.*, puis, dans un second temps, avec ou sans pause significative, est inséré l'adjectif *noir* entre *chat* et *mange*. Nous aurons donc un changement de configurations: l'ensemble *Un chat mange,* d'abord produit en bloc, est ensuite, disjoint par une insertion. Ce qui donne alors une nouvelle configuration de type : *Un chat / noir / mange.*

Ces deux configurations peuvent être simplement représentées de la manière suivante :

- *Un chat mange.* (1 unité),
- *Un chat / noir / mange.* (3 unités),

Où l'insertion de *noir* après *chat* divise la première unité en 3 nouvelles unités.

Nous avons donc cherché à représenter comment, du point de vue de la production, une unité n'est pas *a priori* stable. Rien ne peut prédire, au moment où celle-ci est énoncée, la position qu'elle occupera au sein du texte final. Cette unité y sera-t-elle encore présente ? Ou bien aura-t-elle complètement disparue du texte ? Ou bien encore occupera-t-elle cette même place ? Tout est possible dans cet univers de production où, malgré tous les efforts de planification anticipée, aucune unité ne peut être tenue à l'écart des effets de production.

Cette unité changeante, hautement fonctionnelle, va dorénavant prendre le nom, dans ce travail, de *cellule*. Celle-ci désigne une unité vivante, en plein devenir, au sein d'un ensemble qui est en train de prendre corps, de s'organiser. Ce changement de configuration sera alors représenté visuellement de la manière suivante, à l'aide d'un graphe de trois cellules :

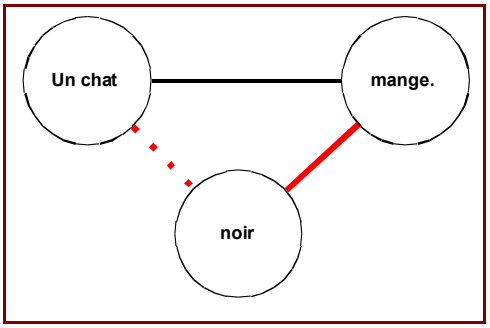

Figure 5. Graphe élémentaire (3 cellules)

Par une analyse dynamique du processus, les trois cellules peuvent être lues de deux manières différentes, et c'est pour cette raison qu'elles doivent être distinctes. Une première lecture, chronologique serait « Un chat / mange / noir », alors qu'une autre lecture, topographique, serait « Un chat / noir / mange. » L'utilisation des graphes permet de conserver simultanément les deux lectures du processus[3].

La première information élémentaire que véhicule une cellule est donc sa position topographique et chronologique ; à cette information élémentaire (identification du jeu originel des cellules de production), il nous est possible d'en rajouter d'autres ; a) comme celle de savoir combien d'*événements* sont inclus, à un moment donné, dans une même et seule cellule ; ou bien encore b) comme celle de savoir si telle ou telle cellule résistera, ou non, aux reformulations de mise en texte pour apparaître dans le texte (final).

Ainsi donc, deux autres informations ont été rajoutées dans le mode de visualisation : la première concerne la taille des cellules qui notera le nombre d'événements produits au clavier ou à la souris : plus une cellule est grosse, plus elle comprend d'événements. La seconde information a trait au jeu de couleurs utilisées (deux ou trois) : les cellules, toutes colorées, peuvent anticiper le texte (final) en distinguant celles qui vont disparaître de celles qui vont subsister. Ainsi, dans l'exemple suivant

[3] Le jeu des traits sera expliqué dans 3.2.

(figure 6), nous proposons un jeu de trois couleurs : les cellules jaunes et rouges sont des *cellules d'ajout*, tandis que les cellules bleues sont des *cellules de suppression*.

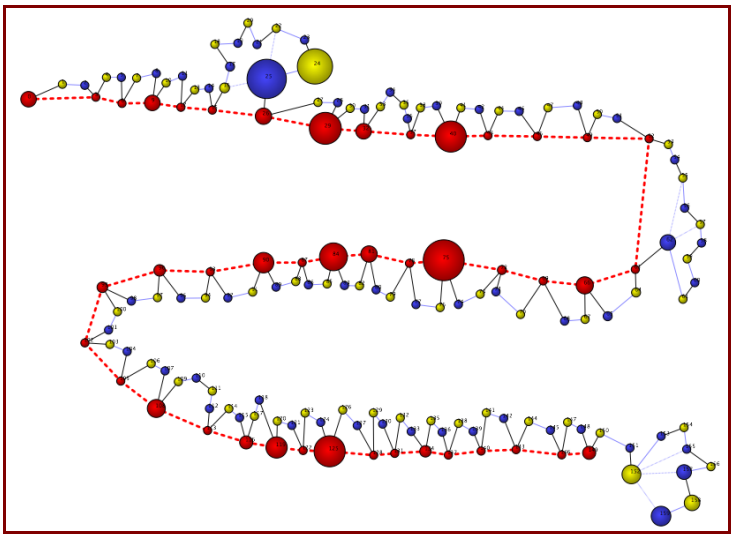

Figure 6. Graphe complet, écriture (15 minutes)

3. 2 Les liens cellulaires

Il est possible d'aller bien plus loin dans la notation des données offertes par les enregistrements, grâce aux liens. Les cellules entretiennent avec d'autres cellules des liens contingents : l'un chronologique, l'autre topographique. Ainsi, un lien chronologique, marquant la succession, entre deux cellules qui ne seront pas supprimées (donc en rouge) sera noté d'un trait gras (le fameux *fil rouge*). Les lignes en trait continu représentant les liens chronologiques, il est donc possible de suivre le processus de manière chronologique en suivant uniquement les lignes continues tandis qu'une lecture topographique se fera par les lignes rouges. Les cellules de suppression portent bien sûr sur une portion de texte qui avait été ajoutée précédemment. Une ligne bleue relie la cellule de suppression à la cellule d'ajout qui lui correspond et cette dernière sera représentée en jaune.

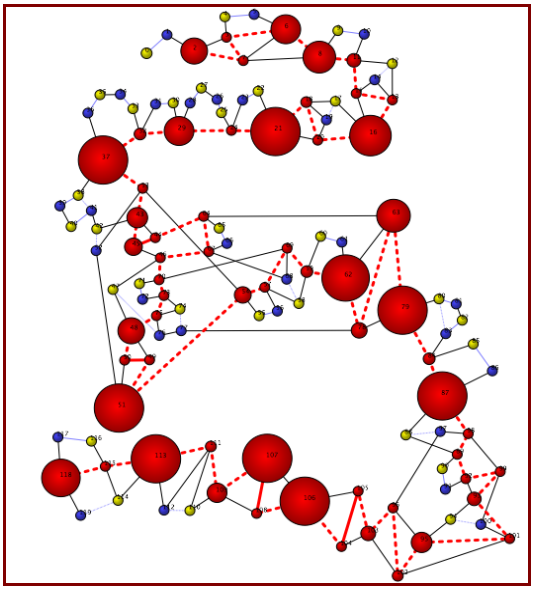

Figure 7. Graphe complet, écriture-réécriture (15 minutes)

La portion centrale du graphe représenté sur la figure 7 (entre les cellules 37 et 79) n'a pas la structure linéaire que l'on trouvait sur la figure 6 (ou que l'on trouve au-dessus et au-dessous). Cette

portion représente un travail plus complexe sur l'avant-texte. Il est aisé de constater que ces liens servent à noter des phénomènes désordonnés nécessaires à la production d'autres phénomènes, plus organisés, qui contribueront à l'accroissement d'un certain ordre, comme la mise en place d'une planification. Il n'est donc pas étonnant de voir de quelle manière l'instabilité cellulaire est potentiellement le lieu et l'occasion de nouvelles productions, mais rien ne garantit, lors de l'énonciation, que ces nouvelles productions vont dans le sens d'une amélioration ou d'une détérioration de ce qui a déjà été écrit (Cogis & Leblay 2010).
De ce point de vue, les grandes représentations dichotomiques proposées par la linguistique génétique sont une aide précieuse. Selon cette approche, l'*écriture à programme* (ou *programmation scénarique*) ne serait qu'un cas particulier de l'*écriture à processus* (ou *structuration rédactionnelle*), l'intention ne pouvant se construire qu'au fur et à mesure de la mise en texte ; ce qui nierait l'idée selon laquelle il existerait une quelconque forme de planification spontanée. Toute planification contiendrait, à ce titre, au moins un seul projet certain : celui de son propre effacement. Toute programmation serait utile, à un moment donné seulement. Elle serait toujours éphémère ; d'où l'idée de recentrer notre travail sur la notion de *dynamisme temporel*, notion générique, commune aux divers champs disciplinaires impliqués dans l'analyse de l'écriture, depuis la génétique du texte, en passant par la psychologie cognitive. Les graphes permettent de montrer le dynamisme temporel en proposant des focales différentes : en pointant le curseur sur une cellule, celle-ci laisse apparaître les événements qui la remplissent.

3.3 Analyse de sous-graphes

Afin d'illustrer précisément notre propos, nous nous proposons maintenant de décomposer certaines opérations d'écriture-réécriture mises en évidence par la génétique du texte. Cette dernière, riche d'une expérience d'une quarantaine d'années, a su, à partir de la technologie du papier-crayon, faire une description précise des brouillons d'écrivains. Récemment, cette approche, principalement développée dans le cadre de la littérature, a trouvé des applications pour la description linguistique de brouillons d'écoliers (Fabre-Cols 1987, 2004 et Doquet-Lacoste 2004) et de brouillons de linguistes (Fenoglio 2012).
A partir des quatre opérations d'écriture décrites par la génétique du texte (Grésillon 1994), nous proposons une modélisation dont les termes sont empruntés à Desbordes (2007), selon la formulation suivante : ainsi, si est considérée une suite initiale d'éléments telle que ABC, il est alors possible de dire de ces éléments qu'ils peuvent *varier* par addition (ABCD), par soustraction (AB), par mutation (ABE) et par permutation (ACB). La manière de décrire la variation au sein d'une *suite*, nous permet de souligner que la notion de cellule ne recouvre pas celle d'opération d'écriture. Si, en effet, elles se recoupent souvent, elles sont associées à des opérations légèrement différentes, les opérations détectées dans le graphe étant purement descriptives alors qu'une composante d'interprétation est présente dans la description des opérations d'écriture.
Dans les deux premiers cas, il s'agit de considérer le jeu élémentaire des cellules : soit elles sont simplement additionnées, soit elles sont soustraites. Voici, à titre d'illustration, des saisies d'écran de différentes visualisations de ce jeu élémentaire de cellules, auquel nous joignons l'opération de mutation (remplacement).

Addition (ajout)

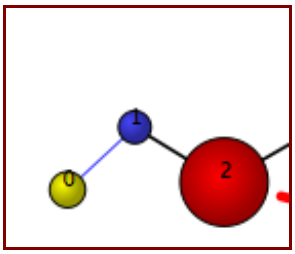

Figure 8. L'addition (ajout)

Dans la figure 8, nous voyons que la cellule 0 a été ajoutée, puis *immédiatement* supprimée par la cellule de suppression 1. Ensuite, la cellule 2 a été ajoutée et elle restera jusqu'à la fin de la production.

Soustraction (suppression)

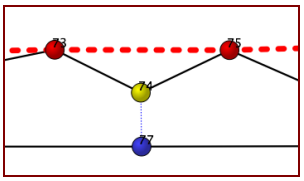

Figure 9. La soustraction (suppression)

Dans l'exemple de la figure 9, les cellules 73,74 et 75 ne faisaient qu'une, jusqu'à ce que la cellule de suppression 77 ne soit associée à une portion de cette cellule, la forçant à être divisée, le texte de la cellule 74 étant supprimé, les textes des cellules 73 et 75 devenant topograhiquement contigus.

Mutation (remplacement)

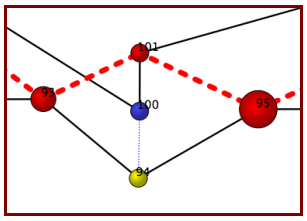

Figure 10. La mutation (remplacement)

La figure 10 représente un exemple de remplacement. Le début du processus est proche de la suppression, la portion de texte correspondant à la cellule 94 étant supprimée par la cellule 100, avant d'être remplacée par le contenu de la cellule 101.

Dans tous les cas de figure, les cellules ne peuvent que se multiplier : toutes les opérations mentionnées ne font qu'augmenter le nombre des cellules présentes dans une écriture. Ajouter des cellules, en supprimer, ou bien même en diviser ne fait que multiplier les cellules.
Pour trouver l'équivalent papier, l'on pourrait dire qu'un scripteur ne peut faire disparaître ce qu'il a écrit qu'en couvrant par des ajouts supplémentaires (ratures, traits) ce qu'il vient de produire ; en agissant ainsi, il ne fait qu'augmenter le nombre de ses propres activités sur la page. Effacer, barrer, biffer sont, en ce sens, toutes des activités de multiplication de l'activité scripturale.
Nous proposons donc d'exploiter ces fichiers d'enregistrement à l'aide d'une visualisation qui a fait ses preuves dans bien d'autres domaines d'application. Cette visualisation par graphes mathématiques offre une grande liberté d'encodage des données en même temps qu'elle offre un confort visuel déterminant pour des applications en recherche sur l'écriture.
D'une manière générale, la visualisation fait intervenir autant que possible des éléments géométriques, le plus souvent colorés : En ce sens, la visualisation est partie prenante de la représentation. L'écriture n'étant pas une activité séparée des autres phénomènes de production, il

ne semble guère possible d'échapper aux tendances contemporaines qui utilisent, très diversement, les visualisations pour rendre compte de la compréhension des phénomènes généraux de production (Okada, Buckingham Shum et Sherborne 2008).

CONCLUSION

La leçon apprise de la méthodologie génétique est bien de mettre en parallèle des écritures, appelées *versions* ou *états*, et donc des représentations de ces productions, sans prétendre à l'excellence de l'un(e) sur l'autre. La démarche que nous avons aussi suivie consiste à passer par des représentations variées de l'expression du temps, sans chercher à en imposer une en particulier qui serait la meilleure possible.

Au titre de représentations modélisées, la génétique propose son choix de transcriptions bipolaires (diplomatique vs linéaire) en réservant une place de choix, le plus souvent malgré elle, à la représentation de l'espace (transcription diplomatique). Comme il a donc fallu représenter *l'espace de l'écriture* en utilisant la transcription la mieux adaptée, il faudrait maintenant apprendre à représenter *le temps de l'écriture* par des représentations, des visualisations qui lui sont le mieux adaptées. Incontestablement, la linéarité (transcription linéaire), notant la succession des données temporelles est la représentation la mieux adaptée. Que dire alors de la linéarité impliquée dans les représentations offertes par les graphes?

D'une manière générale, l'enjeu de ce nouveau genre de représentations linéaires est de permettre une plus grande simplicité, mais aussi une uniformisation des appareils descriptifs des phénomènes linguistique en production. Très concrètement, cela implique une large diffusion des documents numériques produits par des logiciels d'enregistrement : en tout premier lieu, les fichiers d'enregistrement (*logs*) qui restent le document numérique *princeps*, à partir duquel tout se fait. Ceux-ci peuvent être lus et vus 1) *verticalement*, sous la forme de colonnes de chiffres et de lettres, ou bien 2) *horizontalement*, sous la forme de succession de lignes, chaque ligne représentant un événement. En second lieu, les textes linéaires du temps (linear-text) qui imposent *une vision et une lecture uniquement horizontale* et rendent plus familières les données temporelles en évitant une notation chiffrée. Ils permettent de passer d'une *numératie* supportée à une *littératie* acceptée.

La réticence envers la numératie, envers des échelles chiffrées susceptibles d'aider à voir ce qui se passe avec des lettres, est réelle du côté des tenants des sciences humaines. A force d'avoir été contraint à choisir, très tôt, dès l'école, entre lettres ou chiffres, on en vient rapidement à oublier que les deux peuvent aussi se compléter, se représenter mutuellement. Utiliser des outils où apparaissent des suites de chiffres notant, avec précision et exhaustivité, le temps de l'écriture ne devrait surprendre personne : c'est dans notre habitude occidentale de noter le temps en chiffres, depuis la montre (ou le téléphone portable) au calendrier papier ou numérique. Pourquoi alors devrait-on être surpris d'une notation du temps de l'écriture à l'aide de chiffres ? Les outils et les méthodes existent déjà : il suffit de vouloir s'en approcher en acceptant de changer d'habitus d'observation. Adopter une notation du temps de l'écriture en chiffres, c'est faire passer le temps des données lisibles aux données visibles. C'est redonner au temps son expression élémentaire, exhaustive et systématique. Concrètement, c'est refuser d'oublier que l'affichage sur écran, écriture comprise, est fondamentalement construit sur l'enregistrement des données temporelles. Et cette notation des données temporelles en chiffres, le plus souvent invisible pour l'utilisateur ordinaire, n'a pas à l'être pour le chercheur. Mais il reste un important travail à faire dans ce domaine, à commencer par une meilleure fréquentation de ces outils chiffrés que sont les fichiers d'enregistrement. D'aspects souvent rébarbatifs, ils sont facilement mis au rebut ; ils permettent, néanmoins, d'introduire à la temporalité de l'écriture, en associant directement temps et représentation. C'est dans cette direction que nous nous sommes aventurés afin de proposer une

visualisation de ce qu'ils notent en chiffres.
La représentation de la succession et de la durée des opérations d'écriture par l'intermédiaire des graphes est une représentation possible qui vient s'ajouter à celles déjà existantes. Bâtie sur le fichier d'enregistrement, la représentation par graphes permet de visualiser des *patterns temporelles* de certaines opérations d'écriture. Elle rend possible la description de phénomènes *microgénétiques,* centrés sur la *textualisation*, tout autant que la description de phénomènes *macrogénétiques*, centrés sur la *planification*, puisque les graphes offrent dynamiquement la possibilité d'une focale variable.
Des différentes technologies impliquées dans la production écrite, celle de l'écran-clavier donne accès à des informations nouvelles auxquelles celle du papier-crayon ne pouvait donner accès. Notre recherche inaugure, de façon modeste, l'actualité des visualisations par graphe de la dynamique d'écriture, rien d'autre. Il reste un important travail à faire pour les développer et les adapter à des contextes d'écriture variés (Leblay et Caporossi 2014). Ces changements de paradigmes dans la vision de l'avant-texte et la lecture du temps au sein de l'écriture concernent autant la description que l'analyse de phénomènes génétiques, linguistiques, et cognitifs (Alamargot et Lebrave 2010).

BIBLIOGRAPHIE

ALAMARGOT, Denis et LEBRAVE, Jean-Louis (éd.) (2010). « The study of professional writing: A joint contribution from cognitive psychology and genetic criticism », *European Psychologist*, Vol 15(1), 12-22.

ARAGON, Louis (1979). « D'un grand art nouveau, la recherche », HAY, Louis (éd.) *Essais de critique génétique*, Paris, Flammarion, 9.

BENVENISTE, Emile (1966). *Problèmes de linguistique générale*, Paris, Gallimard.

BELLEMIN-NOËL, Jean (1982). « Avant-texte et lecture psychanalytique », HAY, Louis et NAGY, Péter, Avant-texte, texte, après-texte, *Actes du Colloque international de textologie à Mátrafüred* (Hongrie), 13-16 octobre 1978, Akadémiai kiadó, 161-165.

CAPOROSSI, Gilles et LEBLAY, Christophe (2011). « Online Writing Data Representation: a Graph Theory Approach », GAMA, Joao, BRADLEY, Élizabeth et HOLLMEN, Jaakko (éd.), *Lecture Notes in Computer Sciences* 7014, Advances in Intelligent Data Analysis X, Springer Heidelberg, Dordrecht, London, New York, 80-89.

CHANQUOY, Lucile et FENOGLIO, Irène (éd.) (2007). *Langue française* 155, *Avant le texte : les traces de l'élaboration textuelle*, Paris : Larousse.

COGIS, Danièle et LEBLAY, Christophe (2010). « D'une version du texte à l'autre : aperçus sur la morphographie et sa révision en production verbale écrite », LETH ANDERSEN, Hanne, BIRKELUND, Merete, LEBLAY, Christophe et AURIAC-SULARSZYK, Emma (éd.), *Synergie Pays scandinaves* 5, Acquisition et enseignement en production écrite, 65-80.

DESBORDES, Françoise (2007). *Idées grecques et romaines sur le langage*, Paris, ENS editions.

DOQUET-LACOSTE, Claire (2004). *Etude génétique de l'écriture sur traitement de texte d'élèves de Cours Moyen 2, année 1995-96*. Thèse de Doctorat, Lille : Atelier National de Reproduction des Thèses.

FENOGLIO, Irène (éd.) (2012). *Genesis 37*, Le geste linguistique, Paris, PUPS.

FENOGLIO, Irène et ADAM, Jean-Michel (éd.) (2009). *Modèles linguistiques*, Tome XXX, vol. 59, Génétique de la production écrite et linguistique, Edition des dauphins.

FENOGLIO, Irène & GANASCIA, Jean-Gabriel (2008). « Le logicel MEDITE : approche comparative de documents de genèse », CRASSON, Aurèle (éd.) *L'édition du manuscrit. De l'archive de création au scriptorium électronique*, Louvain-la-Neuve, Academia Bruylant, 209-228.

FENOGLIO, Irène et BOUCHERON-PETILLON, Sabine (éd.) (2002). *Langages* 147, *Processus d'écriture et marques linguistiques*, Paris, Larousse.

FABRE-COLS, Claudine (2004). « Les brouillons et l'école : ce qu'a changé la critique génétique », *Le français d'aujourd'hui* 144, 18-24.

FABRE-COLS, Claudine (1987). *Les activités dans les écrits scolaires*. Thèse de Doctorat d'Etat ès lettres, Université Descartes Paris V.

FERRER, Daniel (2008). « La représentation hypertextuelle des manuscrits : quelques leçons de douze années d'expériences », CRASSON, Aurèle (dir.) *L'édition du manuscrit. De l'archive de création au scriptorium électronique*, Louvain-la-Neuve, Academia Bruylant, 189-208.

GRESILLON, Almuth (1994). *Eléments de critique génétique : lire les manuscrits modernes*, Paris, Presses Universitaires de France.

GRESILLON, Almuth et LEBRAVE, Jean-Louis (éd.) (1983). *Langages* 69, Manuscrits – Ecriture. Production linguistique, Paris, Larousse.

GRIZE, Jean-Blaise (1996). *Logique naturelle et communication*, Paris, PUF.

KOLLBERG, Py (1998). *S-notation – a computer based method for studying and representing text composition. Licentiate thesis*, Stockholm, Sweden. Department of Numerical Analysis and Computing Science, Royal Institute of Technology.

LEBLAY, Christophe & CAPOROSSI, Gilles (éd.) (2014). *Temps de l'écriture. Enregistrements et représentations*. Coll. « Sciences du langage, carrefours et points de vue », Louvain-La-Neuve, Academia-Bruylant.

OKADA, Alexandra, BUCKINGHAM SHUM, Simon et SHERBORNE, Tony (éd.) (2008). *Knowledge Cartography. Software Tools and Mapping Techniques,* Springer, London.

PETILLON, Sabine et GANIER, Frank (éd.) (2006). *Langages* 164, *La révision de texte : méthodes, outils et processus*, Paris, Larousse.

PLANE Sylvie, OLIVE, Thierry et ALAMARGOT, Denis (éd.) (2010). *Langages* 177, *Traitement des contraintes de la production d'écrits : aspects linguistiques et psycholinguistiques*, Paris, Larousse.

SEVERINSON-EKLUNDH, Kerstin (1994). « Linear and non-linear strategies in computer-based writing ». *Computers and Composition* 11, IPLab Report 59, Stockholm, 203-216.

STRÖMQVIST, Sven *et al.* (2006). « What Keystroke Logging can Reveal about Writing », SULLIVAN, Kirk P. H. & LINDGREN, Eva (éd.), *Studies in Writing, Vol. 18*, Computer Keystroke Logging : Methods and Applications, Oxford, Elsevier, 45-71.