

**Représentation de la genèse
d'un texte par un graphe**

G. Caporossi
C. Leblay

G-2011-16(F)

Mars 2011

Représentation de la genèse d'un texte par un graphe

Gilles Caporossi

GERAD & HEC Montréal
Montréal (Canada)
gilles.caporossi@gerad.ca

Christophe Leblay

ITEM
Paris (France)
& SOLKI
Jyväskylä (Finlande)
christophe.leblay@kolumbus.fi

Mars 2011

Les Cahiers du GERAD

G-2011-16

Copyright © 2011 GERAD

Résumé

Les outils informatiques de capture en temps réel (Scriptlog, inputlog ou Eye and Pen) procurent des données d'une qualité inégalée (par les autres méthodes) mais très souvent difficiles à exploiter. Ils produisent des séquences extrêmement détaillées décrivant un phénomène complexe (les opérations d'écriture) et il est difficile de saisir l'essence de l'information qu'ils recèlent sans effectuer des traitements préalables. Une approche que nous proposons consiste à agréger les événements en séquences et à travailler sur les liens entre ces séquences. Le déroulement d'une activité d'écriture peut alors se représenter sous la forme d'un graphe dans lequel chaque noeud représente une séquence d'événements alors que chaque arc identifie des relations (spatiales ou temporelles) entre séquences. Cette représentation permet de voir, d'une manière nouvelle, l'activité dans une (certaine) globalité indispensable à l'analyse, ce qui est impossible en travaillant à partir des seules sorties brutes du logiciel. Nous proposons ainsi d'allier la précision des données et une vision globale du processus. Notre méthodologie (alliant génétique du texte et mathématique appliquée) sera exposée et illustrée par des exemples.

Abstract

The computer keystroke logging tools of data capture in real time (Scriptlog, Inputlog, JEdit or Eye and Pen) enable to capture different kind of data which are not, in any case, very easy to exploit. These observational tools produce extremely detailed sequences representing a complex phenomenon (the operations of writing) and it is difficult to grab the essence of information which they conceal without performing prior data treatments. The approach that we propose consists in merging the events in sequences and working on the links between these sequences. The course of an activity of writing can then be represented as a graph in which each node represents a sequence of events whereas each edge identifies spatial or temporal relations between sequences. This presentation allows to see the activity in a new way and draw a global picture of the process necessary to the analysis, what is impossible to do by working from the raw outputs of the software. We offer thus to combine the precision of data and a comprehensive view of the process. Our methodology (combining genetic criticism of the texts and applied mathematics) will be exposed and illustrated by examples.

Introduction

Il existe actuellement plusieurs systèmes permettant de recueillir des données d'écriture en temps réel. Chacun comporte ses particularités depuis *ScriptLog* destiné à l'étude de l'écriture par clavier interposé à *Eye and Pen* qui enregistre les mouvements oculaires durant l'écriture en passant par *InputLog* qui recueille diverses informations dans un environnement naturel. La méthodologie que nous proposons ici est destinée à l'étude des données fournies par *ScriptLog* bien que les concepts utilisés puissent s'appliquer dans d'autres contextes.

Les informations élémentaires que fournit un logiciel tel que *ScriptLog* sont les frappes de touches de clavier ou des clicks de souris. Nous les appellerons *atomes* pour refléter leur aspect insécable et parce que c'est l'unité élémentaire à partir de laquelle tout le texte se construit à la manière d'une molécule complexe.

Schématiquement, on peut dire que *ScriptLog* enregistre les atomes sous la forme d'une liste. Chaque atome y est décrit sur une ligne et caractérisé par le moment où il se produit, sa nature (s'agit-il d'un click de souris ou bien d'une touche frappée?) et la position du curseur dans le texte, comme l'illustre le court extrait sur la Figure 1.

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 1 – Extrait d'un fichier d'enregistrement (log) obtenu à l'aide de ScriptLog

S'il est certain que ces informations permettent d'analyser le processus d'écriture dans ses moindres détails, la finesse de ces détails rend toute analyse directe difficile car le chercheur se voit submergé d'informations qu'il doit classer et regrouper dans le but d'extraire ce qu'il cherche. Cet important volume d'informations rend l'analyse possible seulement pour de courtes, voire de très courtes productions. Le recours à une technique de visualisation du processus d'écriture permet de stimuler l'intuition et l'esprit de synthèse du chercheur. Dans ce sens, certaines tentatives comme *Genèse du texte* (Foucambert, 1992) ont permis de représenter le processus d'écriture et d'aider l'analyse. Ce type de représentation est de fait devenu un standard notamment utilisé dans le logiciel InputLog (Sullivan & Lindgren, 2006, pp 73-94).

Elles reposent sur la création d'un graphique, indiquant la position du curseur en fonction du temps.

Le point faible d'une telle représentation vient du fait qu'au gré des insertions et suppressions dans le texte déjà écrit, la position d'une lettre change. Il est fréquent qu'un caractère écrit à une certaine position apparaisse à une autre position dans le texte final, ce qui montre que la notion de position dans le texte est variable en fonction de l'avancement de l'écriture.

Au lieu de travailler à partir de la position à laquelle se trouvait un caractère lorsqu'il a été produit, nous utilisons un concept plus souple qui consiste à simplement identifier les portions de texte qui sont reliées

soit par un lien spatial (modification, insertion,...) ou temporel. À cette fin, l'outil de modélisation le plus adapté est le graphe au sens de la théorie des graphes plutôt qu'un graphique (dans lequel la position absolue associée à un événement est importante).

Les graphes sont des outils mathématiques basés sur des noeuds ou sommets qui sont éventuellement reliés par des arcs ou arêtes. Si la théorie des graphes est ancienne puisqu'on considère généralement que la théorie des graphes est née avec la présentation du problème des 7 ponts de Königsberg par Euler, au XVIIIème siècle, elle s'est réellement développée au XXème siècle (Berge, 1958, 1970). Les applications de la théorie des graphes sont nombreuses dans le domaine des transports, de la confection d'horaires ou de la communication qui sont autant d'applications naturelles des graphes. Depuis les années 1990, les graphes sont aussi utilisés en sciences humaines par l'entremise des cartes conceptuelles (Novak, 1990). Nous nous proposons ici d'étendre l'utilisation des graphes à la visualisation du processus d'écriture.

1 Description du processus de construction d'un graphe

Les atomes que l'on trouve dans les enregistrements sont de trois types : (i) ajout ou insertion d'un caractère ou d'une espace (ii) suppression d'un caractère ou d'une espace et (iii) déplacement du curseur vers le haut, le bas, la droite, la gauche ou par l'utilisation de la souris¹.

Pour faciliter l'analyse, les séquences d'atomes de même nature (ajouts ou suppressions), contiguës tant au niveau spatial que temporel, sont agrégées en unités fonctionnelles que nous appellerons *cellules* et qui sont représentées par les noeuds d'un graphe. La taille de chaque cellule est un indicateur du nombre d'atomes qui la composent alors que sa couleur indique si la cellule représente un ajout (en jaune ou en clair) ou une suppression (en bleu ou en foncé).

Dans le processus d'écriture, les événements sont reliés par un lien de contiguïté spatiale ou temporelle. Les cellules se construisant l'une après l'autre sont numérotées dans leur ordre de création et reliées par un arc continu indiquant ce lien temporel. Mais, si deux cellules se suivent dans le texte final, elles seront reliées par un arc gras (représenté en rouge quand les couleurs sont disponibles) qui représente le texte final. Il est à noter que les cellules reliées par de tels arcs représentent toutes des ajouts. Le troisième type de lien entre deux cellules est l'appariement entre une cellule de suppression et la (les) cellule(s) représentant l'ajout correspondant. De tels liens sont représentés par un trait fin (bleu quand les couleurs sont disponibles) qui relie une cellule suppression et une cellule ajout, ainsi que le résume le Tableau 1.

Tableau 1 – Représentation des arcs

Arc	gras	normal	fin
Continu	Note une séquence de cellules qui se suivent chronologiquement et spatialement. Apparaît seulement dans le cas d'ajout dans la cellule qui est en train d'être écrite mais pas à la fin de la cellule.	Note une séquence de cellules qui se suivent chronologiquement mais pas spatialement.	Note une séquence de cellules qui se suivent chronologiquement et qui sont liées par une relation de type ajout/suppression. Apparaît lorsqu'est supprimée la cellule qui vient d'être écrite.
Pointillé	Note des cellules qui se suivent dans le texte final mais qui ne se suivent pas chronologiquement.		Note une suppression d'une cellule qui ne vient pas d'être écrite, une suppression dans le texte déjà écrit.


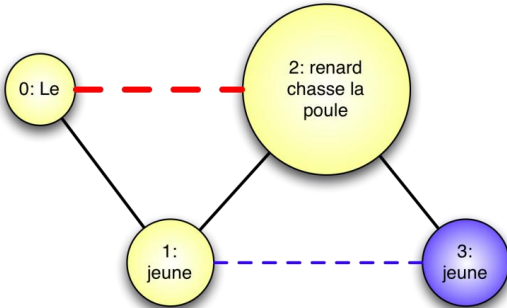
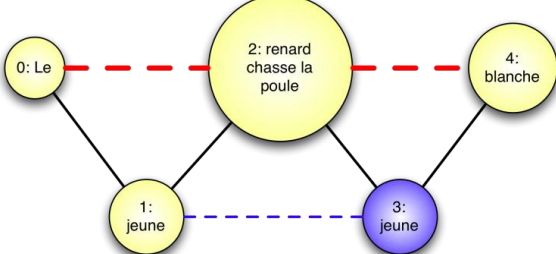
1. Notons qu'il est possible, par l'entremise de la souris, de sélectionner un ensemble de caractères et les effacer simultanément, mais ce type d'événement particulier peut être ramené à une séquence de suppressions successives sans perte de sens au niveau de l'analyse.

Prenons l'exemple d'une personne qui écrit « Le jeune renard chasse la poule » avant de supprimer « jeune », puis d'ajouter « blanche » à la fin d'énoncé (Tableau 2).

Tant que l'auteur ne fait pas de retour sur ce qu'il a écrit, le graphe représentant son travail est composé d'une seule cellule (étape 1).

Après la première insertion, cette cellule « Le jeune renard chasse la poule » est divisée en trois : « Le » d'une part, « jeune » ensuite et « renard chasse la poule » enfin. Le graphe est alors composé de 4 cellules puisque la suppression de « jeune » est une cellule à part entière (étape 2). Enfin l'ajout de « blanche » formera une cellule supplémentaire (étape 3).

Tableau 2 – Description d'un graphe

Étape	Graphe	Texte lisible
1 : Écriture initiale		Le jeune renard chasse la poule
2 : suppression de « jeune »		Le renard chasse la poule
3 : ajout de « blanche »		Le renard chasse la poule blanche

2 Comprendre un graphe

Lire la séquence des cellules sur la trace finale

Tout d'abord, regardons la trace finale, suite de séquences reliées par des traits rouges. Comme les lignes rouges font le lien entre les séquences qui restent dans la trace finale, ce sont toutes des séquences d'ajout

(en jaune ou en clair). Lire le contenu des séquences en suivant le « fil rouge » revient à lire le texte tel qu'il apparaît dans son état final.

La numérotation des cellules suivant l'ordre chronologique, l'ordre dans lequel apparaissent les cellules de la trace finale nous indique la manière dont le texte a été écrit. Si les numéros apparaissent dans un ordre croissant tout au long de la trace finale, c'est que l'auteur n'a pas fait de retours dans le texte déjà écrit qui soient restés dans la version finale du texte.

Selon cette manière de regarder le graphe, on peut identifier des ensembles de séquences que nous qualifierons de linéaires si les cellules de la trace finale sont dans un ordre croissant et que les cellules qui ne sont pas sur la trace finale (séquences qui ont été supprimées durant le processus) forment aussi une chaîne croissante comme le représente la Figure 2. Il est à noter que la forme du graphe, c'est-à-dire la forme obtenue par l'agencement des traits et des noeuds-cellules, n'as pas d'incidence sur la valeur de la représentation : ce qui compte, c'est que soient reliées, ou non, deux cellules.

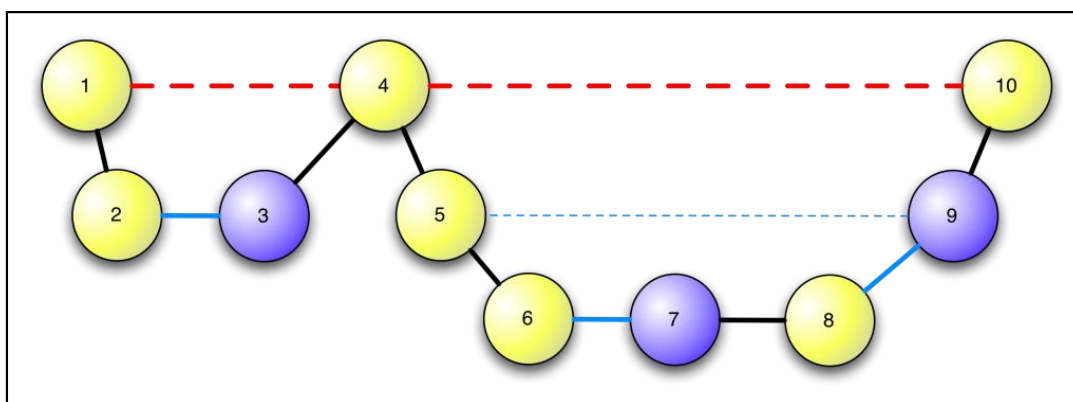


Figure 2 – Représentation d'une séquence linéaire

Les articulations

Tout d'abord, intéressons nous à une cellule qui, si elle était enlevée du graphe, le déconnecterait. Il apparaît qu'après avoir ajouté cette cellule, le scripteur n'est pas revenu sur les zones du texte précédentes. Nous appellerons une telle cellule *articulation*. Remarquons que si une cellule de la trace finale d'un ensemble linéaire est une articulation, toutes les autres cellules de la trace finale de cet ensemble linéaire le seront aussi. Les articulations qui sont les plus intéressantes à étudier se situent en début et fin d'ensemble linéaire.

Identifier des zones d'écriture retravaillées

Un autre phénomène intéressant est celui des « zones d'écriture ». On voit dans certains cas que le scripteur a écrit un passage, puis a retravaillé ce même passage un peu plus tard. Le trait épais (rouge) qui représente l'enchaînement des séquences qui restent dans la version finale du texte est alors « encadrée » par deux séquences temporelles en trait continu puisque la portion de texte correspondante a été construite en deux phases successives. Un exemple de double passage est représenté sur la Figure 3.

3 Exemple d'analyse de graphe

Le graphe suivant représente le processus d'écriture d'un court texte de 15 minutes, emprunté à un corpus enregistré sous ScriptLog 1.4 (version Mac), uniquement composé de scripteurs finnophones (Leblay, 2009). Ce corpus intègre des scripteurs novices (première année de l'université finlandaise, langue française,

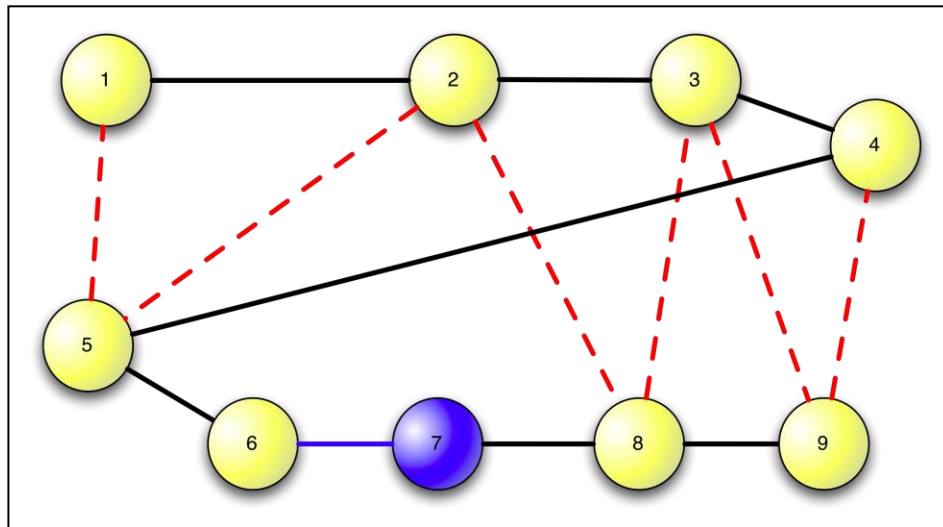


Figure 3 – Représentation d’une portion de texte travaillée en deux passages

matière principale) et des scripteurs experts (enseignants universitaires)². Le texte ci-dessous présenté est une production experte. En voici, pour commencer, le texte produit, achevé, dans son format *.txt*, donné par le logiciel :

J’aimerais vivre, comme la plupart des Finlandais, à la campagne près d’un lac mais pas très loin de la ville. Cette réponse est pourtant très banale, non-imaginaire, ou sans imagination. Sans aucune limite d’imagination, en quinze minutes, ce n’est pas très facile à expliquer.

Ce qui est très important pour moi, c’est la nature et mon chat. Donc, il faut qu’il y ait le ciel bleu avec les nuages et la mer, et mon chat. D’autre part, j’aime bien les films science-fiction et le monde qu’on y décrit sauf la manque de la nature - la nature est toujours polluée ou même inexistante. Comme dans le Fifth Element. Dans ce film, le décor était bien fait et cela semble un environnement très réaliste dans quelques siècles. Les créatures également m’ont beaucoup plus. Il s’agit des extra-terrestres qui ont des traits des animaux, des insectes ou autres. Peut-être est-ce l’envie de traiter les animaux de manière aussi importante que les êtres humains, les hommes.

Finalement, outre la nature, j’aimerais qu’il n’y ait pas de guerres, c’est-à-dire que les hommes et les autres êtres vivants puissent vivre paisiblement ensemble. Je ne suis pas trop imaginative comme je n’ai pas pensé auparavant à cette thématique. Pour moi, l’idéal, c’est la tranquillité, la beauté.

Afin de mieux saisir l’intérêt d’une représentation dynamique par graphe, nous en donnons, au préalable, la transcription linéaire (au format *.doc*) du même texte, où sont notés en gras les retours dans le *déjà écrit*³.

2. La tâche choisie a permis de prendre le scripteur *sur le vif*, au sein de son inquiétude en cherchant de saisir un état de ce qui est en mouvement. Elle a été, en cela, limitée à 15 minutes afin de prendre un simple échantillonnage de la production de chaque scripteur. La tâche d’écriture a été la suivante : *Décrivez votre environnement idéal ainsi que les personnes ou les êtres qui pourraient y vivre. Imaginez-le sans vous limiter. Vous pouvez écrire à la première personne (Je), ou à la deuxième (Tu), ou bien encore à la troisième (Il/Elle).*

3. La sémiologie de ces conventions se veut simple : quatre caractères sont pris en compte : le simple, le ~~barré~~, l’italique et le **gras**, en cherchant à associer, autant que possible, un caractère à un geste. Ces conventions appellent quelques remarques relatives aux gestes. Tout d’abord, au geste d’ajouter est associé le caractère sans variation, le caractère simple de la police. Ensuite, au geste multiple de supprimer est associé le caractère barré ; si cette suppression est immédiate (ce qui est, de loin, le cas le plus fréquent), le caractère barré est doublé de celui de l’italique. Mais s’il s’agit d’une suppression différée, seul le caractère de l’italique est pris en compte. Enfin, les gestes de remplacer et de déplacer, ne font que doubler ceux d’ajouter et de supprimer : ils sont le simple reflet des opérations composées, et ne peuvent donc prétendre à des caractères propres, indépendamment de ceux déjà utilisés.

LIEU IDÉAL Lieu idéal

J'aimerais vivre,¹ c²omme la plupart des Finlandais,³ à la campagne près d'un lac mais ~~t~~pas très loin ~~d~~e⁴ la ville⁵. Cette réponse est pourtant très banale, ~~n~~io⁶n-imaginaire⁷, ou sans imagination. Sans aucune limite d'imagination, en quinze minutes⁸, ce n'est ~~a~~/~~t~~ pas très facile à expliquer. ~~Au fait~~ Ce qui est tr⁹ès ¹⁰important pour moi, c'est la nature et mon chat. Donc, il faut qu¹¹il y ait le ciel bleu **avec les nuages et** la mer¹², **et mon chat** ~~hat~~7./~~Da~~¹³~~ut~~¹⁴~~re~~¹⁵ **autre**¹⁶ **part**, j'¹⁷ ~~a~~¹⁸ aime bien les films science¹⁹-fiction²⁰ et le monde qu'on y décrit sauf la manque de la nature – la nature est toujours polluée ou même inexistante.²¹ Comme dans le Fif²²th element. Dans ~~je~~ ~~film~~ ce film, le décor était bien fait et cela semble **un environnement**²³ très réaliste dans quelques siècles²⁴. Les créatures également m'ont beaucoup plus. Il s'agit des extra-te²⁵-~~rr~~²⁶restres qui ~~ont~~ des traits des animaux, des insectes ou autres. Peut-être est-ce l'envie de ~~voir~~²⁷~~traiter~~²⁸ les ani~~maux~~²⁹ ~~maux~~ **de manière**³⁰ aussi important³¹se que les êtres humains, **les hommes**³².³³

Finalement, outr³⁴e la nature³⁵, j'aimerais qu'il n'y ait pas de guerres, c'est-à-dire **que**³⁶ les hommes et les autres êtres vivants puissent vivre paisiblement ensemble³⁷. Je ne suis pas trop³⁸ imaginative³⁹ comme je n'ai pas pensé auparavant à vette thématique. Pour mois⁴⁰, l'⁴¹idéal, c'est la tranquillité, la beauté.⁴²

Ce même processus d'écriture est maintenant représenté sous la forme d'un graphe comme on le voit sur la Figure 4. Dans ce graphe, apparaissent 119 cellules. S'il donne, à premier abord, l'impression d'être relativement linéaire, un examen plus détaillé nous indique qu'il peut se décomposer en 5 zones d'écriture consécutives, articulées par les cellules 37, 79, 87 et 103.

Zone d'écriture 1 (cellules 0-37)

La portion entre les cellules 0 et 37, représentée ci-dessous (Figure 6), est relativement linéaire et nous pourrions profiter de cette apparente simplicité pour identifier les opérations qui la composent.

À titre de comparaison, la Figure 5 représente la transcription linéarisée correspondant aux cellules 0-37 alors que le graphe est représenté sur la Figure 6.

L'écriture est très précise : La cellule 0 [LIEU IDÉAL] a aussitôt été supprimée, puis les cellules 2 [Lieu idéal<CR><CR>J'aimerais vivre] à 4[i] ont été ajoutées avant que la cellule 4 ne soit supprimée à son tour. La cellule 6 [omme la plupart des Finlandais,] a alors été ajoutée à la fin du texte avant que la cellule 7 [,] ne soit insérée entre la cellule 2 et la cellule 3 [c]. Les cellules 8 [à la campagne près d'un lac mais] et 9 [t'] ont alors été ajoutées avant que cette dernière ne soit supprimée. Sans détailler ainsi tout le processus d'écriture, remarquons que les cellules 11 [pas très loin d], 12 [a] et 13 [la ville] ont été ajoutées avant que la cellule 12 ne soit supprimée pour être remplacée par la cellule 15 [e]. Un autre remplacement a lieu juste après alors que la cellule 17 [i] est remplacée par la cellule 20 [o]. De la cellule 21 [, ou sans imagination. Sans aucune limite d'imagination, en quinze minutes] à la cellule 37 [ès important pour moi, c'est la nature et mon chat. Donc, il faut qu], on remarque simplement des corrections à la fin du texte qui vient d'être écrit.

Zone d'écriture 2 (cellules 37-79)

La zone deux comporte les cellules 37 jusqu'à 79. Voici, dans un premier temps, une transcription linéaire, suivie du graphe équivalent (Figure 8).

La linéarité du processus se rompt entre les cellules 37 et 79 comme on le voit sur la Figure 8⁴.

4. Détail de l'analyse de la construction de la partie centrale du texte représenté sur la Figure 6 : On peut d'abord constater que ce qui a été écrit entre les séquences 37 et 43 ont toutes été supprimées immédiatement et que la séquence 53 sera insérée plus tard entre le noeud 37 et le noeud 43. Sans aller dans les détails, on peut rapidement remarquer que la portion 37-79 a

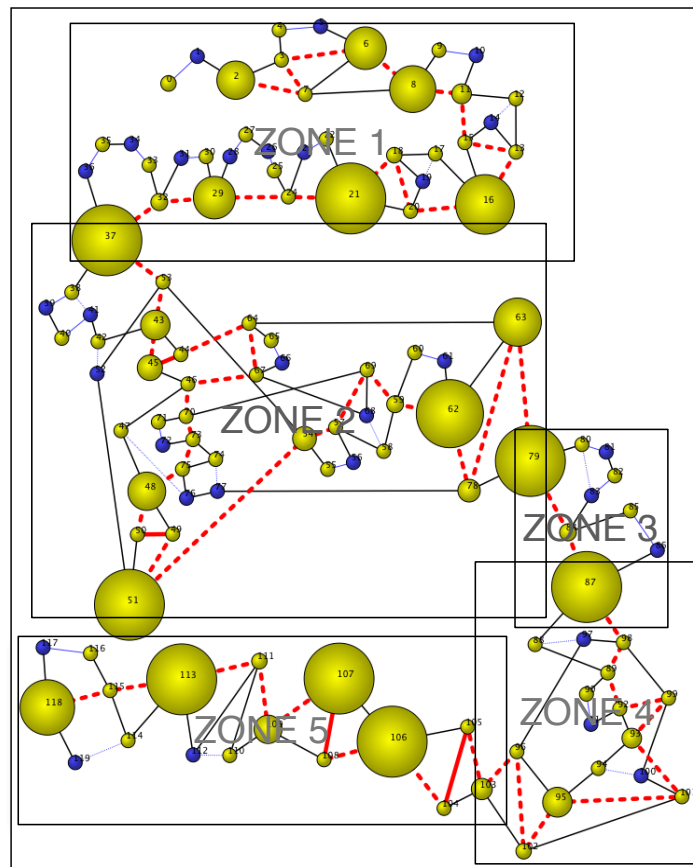


Figure 4 – Graphe représentant l'écriture du texte

LIEU IDÉAL Lieu idéal

J'aimerais vivre,¹ c~~o~~mm²e la plupart des Finlandais,¹ à la campagne près d'un lac mais ~~n~~pas très loin ~~d~~ae² la ville². Cette réponse est pourtant très banale, ~~n~~io³n-imaginaire³, ou sans imagination. Sans aucune limite d'imagination, en quinze minutes-, ce n'esa/~~t~~ pas très facile à expliquer.

~~Au fait~~ Ce qui est tr-ès-ès important pour moi, c'est la nature et mon chat. Donc, il faut qu

Figure 5 – Transcription correspondant aux cellules 0-37

été beaucoup plus travaillée que la portion initiale. Si on désirait une analyse poussée du processus, on pourrait regarder les différentes zones de ce graphe. Par exemple, les séquences 46-47 et 48 ont été travaillées de manière particulière puisque 70, 71 (supprimée aussitôt) et 73, 74 et 75 sont insérées entre 46 et 47, avant que la séquence 47 ne soit supprimée, ce qui revient à dire que les séquences 70,73,74 et 75 remplacent la séquence 47. La séquence 74 sera ensuite supprimée à son tour. Sur cette portion, la séquence des cellules sur la trace finale est 37-53-43-45-44-64-67-46-70-73-75-48-50-49-51-54-57-69-59-62-78-63-79. Les numéros de cette portion laissent voir plusieurs campagnes d'écriture et le graphe montre que la séquence 51 a été suivie de la suppression de 42, puis insertion de 53 qui remplace 42. Viennent ensuite les séquences 43-45-44 qui sont linéaires, à l'exception d'une insertion dans la dernière séquence (45 est insérée entre 43 et 44) suivie de 64-67 qui est insérée entre 44 et 46. Par la suite, 70-73 et 75 remplacent la séquence 45. Entre ces deux modifications, l'auteur a remplacé la séquence 58 par la séquence 69. Pour prendre l'ordre chronologique de ce travail, et si on néglige les corrections mineures dans le texte qui vient d'être écrit, on dirait que le scripteur a inséré 64 et 67 entre 44 et 46, puis qu'il a remplacé 58 par 69 avant de revenir sur remplacer 47 par 70-73-74 et 75, à ce moment là, il supprime 74. Puis il continue son travail en insérant la cellule 78 entre 62 et 63, puis 79 est ajouté à la fin du texte.

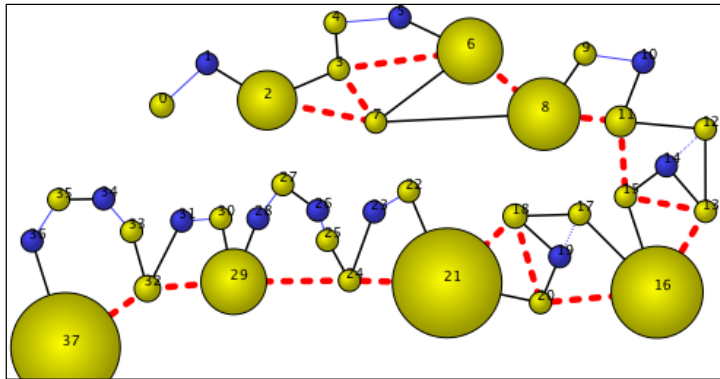


Figure 6 – Zone d'écriture 1

ès important pour moi, c'est la nature et mon chat. Donc, il faut qu'il y ait le ciel bleu **avec les nuages** et⁴ la mer^{←⁴→}, et mon **chat**-hat⁷./De^{autre} autree¹¹ **part**, j'⁹ j'¹⁰ aime bien les films science⁵-fiction^{←⁵→} et le monde qu'on y décrit sauf la manque de la nature – la nature est toujours polluée ou même inexistante.[↑][→]⁶[↓] Comme dans le Fifth element. Dans ~~je~~-film ce film, le décor était bien fait et cela semble **un environnement**¹² très réaliste dans quelques siècles[↖]⁷[↗]⁰[↖]⁸[↗]⁹→¹⁰[↖]¹¹[↗]¹²[↖]. Les créatures également m'ont beaucoup plus. Il s'agit des extra-te

Figure 7 – Transcription qui correspond aux cellules 37-79

Nous procéderons ici à un autre type d'analyse plus général qui consiste à identifier quand chaque cellule constituant la trace finale a été ajoutée. La séquence de cellules que l'on retrouve dans le texte final est la suivante en suivant le trait épais (en rouge) : 37-53-43-45-44-64-67-46-70-73-75-48-50-49-51-54-57-69-59-62-78-63-79.

Pour identifier les *passages de travail* dans cette séquence, nous listons les valeurs des cellules de chaque passage de travail. Lors d'un passage donné, la séquence des valeurs est toujours approximativement croissante, ce qui signifie que d'une cellule à la suivante, les valeurs doivent être croissantes ou avoir une différence d'au plus 1 (dans le cas où il y aurait une insertion dans la séquence qui est en train d'être écrite, qui se reconnaît au trait continu épais), comme on peut le voir sur le graphe de la Figure 8.

De fait, une telle séquence ne peut comporter de valeurs supérieures à une valeur non prise en compte. Ici, le premier passage est composé des cellules 37-43-45-44-46-48-50-49-51. Les cellules suivantes ne font pas partie de ce passage car elles ont des valeurs supérieures à 53 qui a été sauté au début de la séquence (elles ont donc été ajoutées lors d'un passage ultérieur). Le second passage a permis d'ajouter les cellules 53-54-57-59-62-63. La plus petite valeur non prise en compte est alors 64. Les cellules 64-67-69 ont été ajoutées lors du troisième passage après lequel la plus petite valeur sautée est 70. Les cellules 70-73-75-78-79 ont été ajoutées lors du quatrième et dernier passage. Cette analyse des séquences nous indique que les traces qui resteront dans le texte final ont été ajoutées lors de 4 passages successifs, ce qui explique la complexité du graphe correspondant à cette portion de texte.

Zone d'écriture 3 (cellules 79-87)

La troisième zone d'écriture est très brève, comme le montre la transcription sur la Figure 9.

La brièveté de l'écriture fait apparaître un travail très linéaire du scripteur (voir Figure 10).

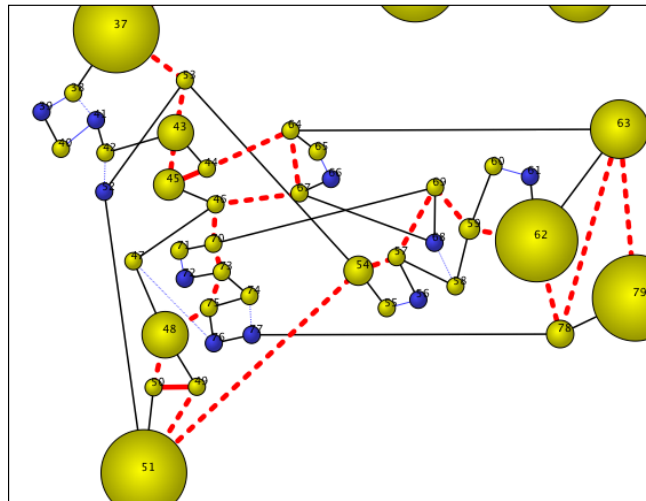


Figure 8 – Zone d'écriture 2

. Les créatures également m'ont beaucoup plus. Il s'agit des extra-terrestres
qui ont des traits des animaux, des insectes ou autres. Peut-être est-ce l'envie de

Figure 9 – Transcription qui correspond aux cellules 79-87

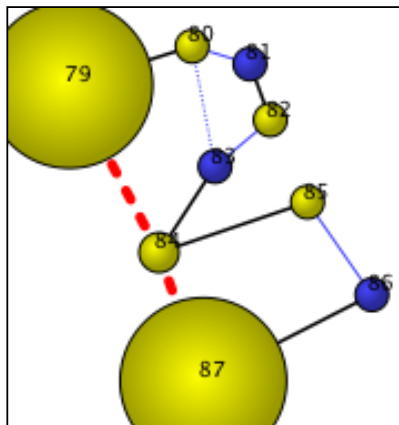


Figure 10 – Zone d'écriture 3

La cellule 79 est caractéristique de ce que nous appelons une *articulation* puisque le scripteur ne reviendra jamais sur le texte déjà écrit jusque là. Entre les cellules 79 et 87, le travail est très linéaire, mais cette linéarité se brise à nouveau à partir du cellule 87 qui est un autre point d'articulation important, comme l'illustre bien la figure précédente (10).

Zone d'écriture 4 (cellules 87-103)

Entre des cellules 87 et 103, on remarque très nettement une correction linéaire du texte déjà écrit. Sur la Figure 12, on remarque que la trace finale du texte oscille entre les trames linéaires 87-95 et 96-103, comme si cette trace produite était au centre de la double hélice qui représente une molécule d'ADN. Remarquons

94 - Ajout	s	ont des traits des animaux, des insectes ou autres. Peut-être est-ce l'envie de voir les animaux aussi importants
95 - Ajout	que les êtres humains	ont des traits des animaux, des insectes ou autres. Peut-être est-ce l'envie de voir les animaux aussi importants que les êtres humains
96 - Ajout	.	ont des traits des animaux, des insectes ou autres. Peut-être est-ce l'envie de voir les animaux aussi importants que les êtres humains.
97 - Suppression - 88	voir	ont des traits des animaux, des insectes ou autres. Peut-être est-ce l'envie de les animaux aussi importants que les êtres humains.
98 - Insertion (remplacement)	traiter	ont des traits des animaux, des insectes ou autres. Peut-être est-ce l'envie de traiter les animaux aussi importants que les êtres humains.
99 - Insertion	de manière	ont des traits des animaux, des insectes ou autres. Peut-être est-ce l'envie de traiter les animaux de manière aussi importants que les êtres humains.
100 - Suppression - 94		ont des traits des animaux, des insectes ou autres. Peut-être est-ce l'envie de traiter les animaux de manière aussi important que les êtres humains.
101 - Insertion (remplacement)	e	ont des traits des animaux, des insectes ou autres. Peut-être est-ce l'envie de traiter les animaux de manière aussi importante que les êtres humains.
102 - Insertion	, les hommes	ont des traits des animaux, des insectes ou autres. Peut-être est-ce l'envie de traiter les animaux de manière aussi importante que les êtres humains, les hommes.

Zone d'écriture 5 (cellules 103-119)

La dernière zone d'écriture est présentée d'abord sous sa forme transcrite en Figure 13.

La portion finale du texte représenté sur la Figure 14 s'est réalisée d'une manière quasiment linéaire, avec des simples insertions dans la dernière séquence, des suppressions immédiates et deux suppressions différées : les cellules 110 et 114 ont respectivement été supprimées en 112 et 119, comme le montrent les arcs en pointillé (bleu).

4 Représentation des opérations d'écriture

Les quatre opérations d'écriture (ajout, suppression, remplacement et déplacement) sont un emprunt de la génétique du texte, en quête d'outils, à la linguistique structurale. Comme le souligne Grésillon (1994, 150), dans ce qui relève d'un mouvement de transfert notionnel, la description de la *substitution* a changé de statut en passant de la *structure* à la *genèse* : « *Initialement symétrique et a-chronique, la substitution appliquée à l'analyse des manuscrits est munie de la dimension du temps.* » C'est précisément cette dernière dimension du temps qui, conjugué à la dimension de l'espace, permet de penser autrement ces opérations d'écriture en leur offrant une autre manière d'être vues et analysées.

Afin de terminer ce travail, il nous a semblé intéressant de pouvoir présenter ce que peuvent être des patterns des opérations d'ajout vs d'insertion, des opérations de suppression (immédiate vs différée), et de l'opération de remplacement.

Finalement, outr¹⁷e la nature $\leftarrow^{17}\rightarrow$, j'aimerais qu'il n'y ait pas de guerres, c'est-à-dire **que**¹⁸ les hommes et les autres êtres vivants puissent vivre paisiblement ensembl¹⁸. Je ne suis pas trop¹⁹ imaginative $\leftarrow^{19}\rightarrow$ comme je n'ai pas pensé auparavant à vettte thématique. Pour mois²⁰, l'édéal, c'est la tranquillité, la beauté.²⁰

Figure 13 – Transcription correspondant aux cellules 103-119

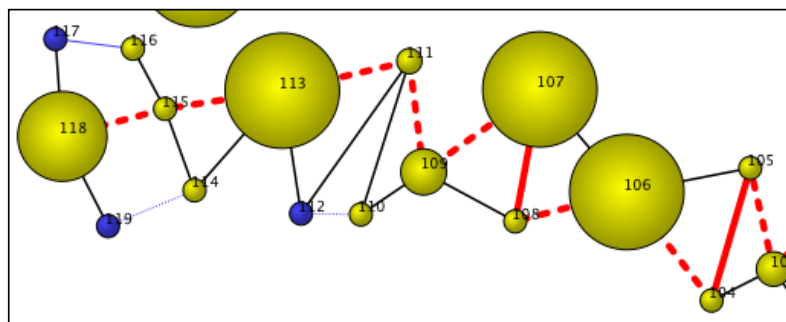


Figure 14 – Zone d'écriture 5

Voyons d'abord, les opérations d'*ajout* et d'*insertion*. Celles-ci entrent souvent en concurrence, et l'apport des représentations par graphe pourrait aider à mieux les voir.

Ajout et insertion

Il convient de réserver le terme d'*ajout* au fait d'ajouter par retour dans le déjà écrit et de nommer *inscription* ce qui relève du fait d'ajouter à la suite du déjà écrit. L'usage du terme d'*inscription* est devenu quasi indispensable, non seulement, comme nous venons de le voir par opposition à celui d'*ajout*, mais aussi pour distinguer le niveau purement graphique du niveau linguistique; nous avons alors la configuration suivante :

- Niveau graphique : où les termes d'*inscription* (au sens d'ajouter à la suite de) et d'*insertion* (ajouter à l'intérieur de) se complètent ;
- Niveau linguistique : où apparaissent les quatre opérations génétiques (ajouts, suppression, remplacement et déplacement).

Les termes d'*inscription* et d'*insertion* entrent donc en système avec celui d'*ajout*. Dans notre travail, l'opération d'*inscription* prendra deux formes différentes

En tant que tel, l'*inscription* à la fin du texte écrit ne se représente que par la cellule claire (en jaune) qui conclut l'enregistrement. Sur le graphe, tous les cellules jaunes correspondent à des *ajouts* ou des *inscriptions*. Selon le cas, ces ajouts restent jusqu'à la fin et apparaissent selon le fil rouge du graphe.

D'un point de vue représentation, l'ajout sera différent selon qu'il intervienne dans la cellule en train d'être écrite ou dans une cellule déjà écrite. Le premier cas de figure est représenté sur la Figure 15 qui illustre l'écriture des mots « le chat » entre lesquels l'adjectif « beau » serait ajouté juste avant l'écriture du mot « chat ». L'ajout, dans une séquence en train d'être écrite, est aisé à identifier puisque c'est le seul cas où une ligne continue rouge apparaît; cette ligne indique que les séquences se suivent du point de vue chronologique et spatial bien que leur ordre soit inversé. Si les deux cellules consécutives étaient dans le même ordre, elles ne formeraient qu'une, par définition.

L'ajout différé est représenté sur la Figure 16 alors que l'adjectif « beau » est ajouté après que d'autres cellules aient été ajoutées au texte, par exemple, un mot supplémentaire effacé ensuite. Ici, l'auteur a écrit « le chat gris », a effacé « gris » puis a inséré « beau » avant « chat ».

Suppressions immédiate et différée

La suppression peut se caractériser par des sous graphes différents selon qu'elle intervient juste après l'écriture ou si elle intervient après qu'une autre cellule ait été ajoutée. Le cas d'une suppression immédiate (par exemple après une

erreur de frappe au clavier) est représenté sur la Figure 18 alors qu'une suppression différée se traduira par le sous graphe représenté sur la Figure 17. Les cellules de suppression sont représentées en sombre (en bleu) sur le graphe et sont associées aux séquences d'ajout, en clair (en jaune), qu'elles ont éliminées.

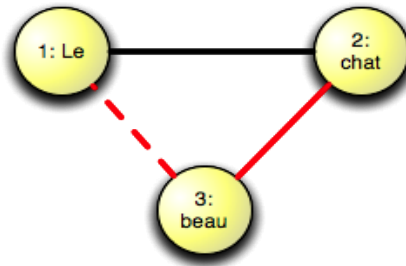


Figure 15 – Ajout

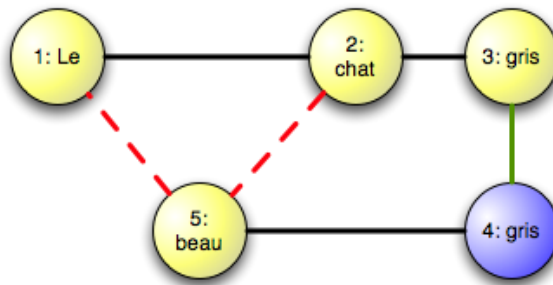


Figure 16 – Insertion différée

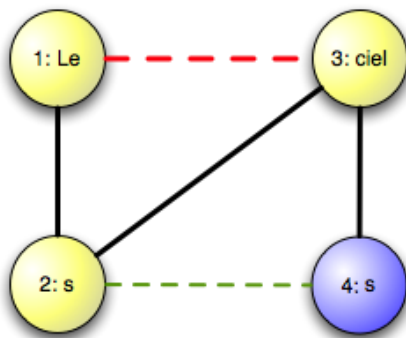


Figure 17 – Suppression différée

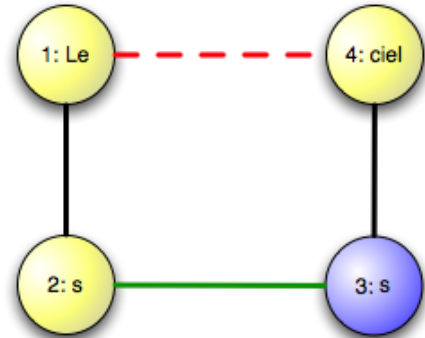


Figure 18 – Suppression immédiate

Remplacement (et déplacement)

En plus de ces opérations simples, il existe des opérations composées qui peuvent être considérées comme des séquences d'opérations simples. Citons le déplacement de texte et la substitution.

Si d'un point de vue purement technique, on peut considérer le remplacement comme une suppression suivie d'un ajout à la même place, il est difficile d'affirmer que cette séquence constitue nécessairement un remplacement dans l'esprit de l'auteur. De même, nous pouvons représenter une opération complexe par un sous graphe caractéristique mais il est difficile d'affirmer que le sous graphe correspond toujours à la même opération. L'identification de sous graphes pour trouver certaines opérations est donc une étape utile mais ne comportant aucune garantie. Prenons l'exemple d'un auteur qui écrirait « le beau chat » avant de remplacer « beau » par « petit ». Cette séquence est composée d'un ajout initial suivi du remplacement de « beau » par « petit », comme l'illustre la Figure 19.

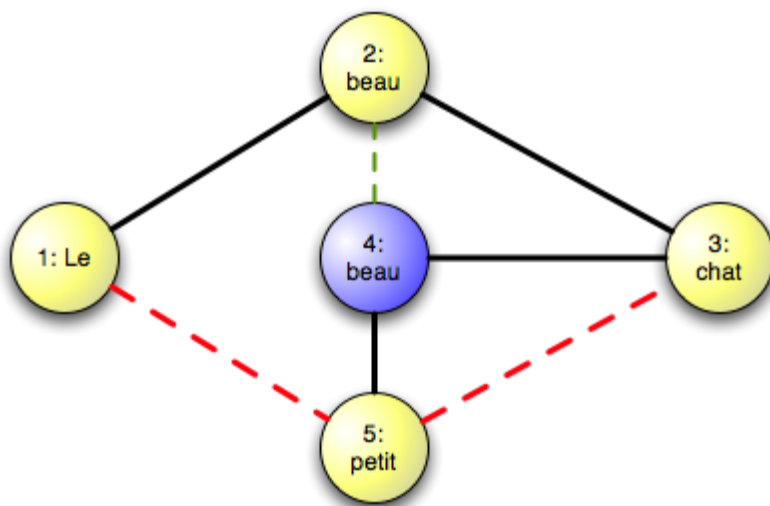


Figure 19 – le remplacement

Ce cas de figure illustre un remplacement au sein de la cellule qui vient d'être écrite, mais pas à la fin de cette dernière. Le remplacement peut prendre deux autres formes : remplacer la fin de la dernière cellule écrite ne se distingue pas de la suppression à la fin de la dernière cellule écrite qu'on remarque sur la Figure 18. Le remplacement peut aussi se faire dans une autre séquence, plus ancienne, en quel cas le sous graphe correspondant sera très similaire à celui de la Figure 17 à l'exception que le trait de lien chronologique entre la cellule 3 et 4 qui serait absent.

Il est à noter que les sous graphes peuvent varier selon ce qui se passe juste avant ou juste après l'opération. Par exemple, la Figure 18 serait différente si l'auteur n'avait pas ajouté de texte juste après avoir supprimé « s ».

Le cas du déplacement est techniquement plus facile à identifier puisqu'il correspond à la suppression d'une cellule juste avant une insertion de la *même* cellule à un autre endroit. Identifier un déplacement nécessite donc une lecture des cellules concernées. Le cas des déplacements désynchronisés est plus difficile à identifier puisque la cellule supprimée ne sera ajoutée que plus tard ; il est difficile d'identifier un lien entre ces deux cellules à moins de tenir compte de la mémoire de l'ordinateur dans le cas où la fonction « couper/coller » serait utilisée.

5 Conclusion

Si la génétique du texte (du brouillon sur papier) n'a pas eu encore les moyens techniques de mettre systématiquement le temps de l'écriture en évidence, elle ne s'est pas privée de commenter ce manque en réclamant souvent l'accès à des outils issus des recherches quantitatives.

L'écriture produite sur écran donne enfin accès à son temps, si celui-ci est enregistré à l'aide d'un logiciel de *temps réel* : le geste d'écriture sur écran, différent de celui qui se produit sur papier, se construit néanmoins de la même manière. Le recours aux opérations d'écriture est un invariant auquel aucun scripteur n'échappe, dans sa production écrite.

La génétique, bâtie à partir des brouillons écrits sur du papier, a déjà entrepris de parler du temps et ses gloses verbales sont bien connues : mais, il lui est parfois difficile d'en dire plus, sans glisser dans l'interprétation pure et simple.

Pour passer du côté de la visibilité, il faut des outils d'écriture différents : la réticence envers des échelles chiffrées susceptibles d'aider à voir ce qui se passe avec des lettres, est encore réelle du côté des tenants

des sciences humaines. Adopter une notation du temps de l'écriture en chiffres, c'est redonner au temps son expression élémentaire, exhaustive et systématique. Concrètement, c'est comprendre 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 ou pour l'enseignant dans le domaine de l'écriture.

Au titre de représentations déjà modélisées, la génétique propose déjà son choix de transcriptions bipolaires (*diplomatique* vs *linéaire*) en réservant une place de choix, exclusive, 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 faut maintenant apprendre à représenter l'espace de l'écriture associé au temps de l'écriture par des représentations, des visualisations adaptées.

La démarche consisterait donc à passer par des représentations qui, bien que construites nécessairement sur des chiffres imposés par le temps de la production écrite, puisse représenter la dynamique de l'écriture dans sa double dimension, dans sa dimension spatio-temporelle.

D'une manière générale, l'enjeu de ces représentations par graphe est de permettre une plus grande simplicité, mais aussi une *uniformisation* des appareils descriptifs. Très concrètement, cela impliquerait une diffusion des documents numériques produits par des logiciels d'enregistrement : à partir du fichier d'enregistrement (*log*) qui reste le *document numérique princeps*, il est dorénavant possible de représenter la genèse par des graphes, où sont notés *simultanément* les liens spatiaux et les liens temporels (succession et durée) des opérations d'écriture. Cette approche permet de mettre en vue, pour les analyser, les *zones d'écriture* distinctes ainsi que de pouvoir proposer de *patterns* simples des opérations d'écriture.

Il s'agit d'une description exhaustive et systématique autant des phénomènes *microgénétiques* que *macrogénétiques* puisque les graphes offrent dynamiquement la possibilité d'une focale variable.

Références

- [1] Berge, C. 1958. *Théorie des graphes et ses applications*. Paris : Dunod.
- [2] Berge, C. 1970. *Graphes et hypergraphes*. Paris : Dunod.
- [3] Foucambert, J. 1992. La recherche *Genèse du texte*, *Les Actes de lecture* 37, 73-76.
- [4] Grésillon, A. 1994. *Éléments de critique génétique : lire les manuscrits modernes*. Paris : Presses universitaires de France.
- [5] Leblay, C. 2009. Les invariants processuels. En deçà du bien et du mal écrire. *Ecrits de savoir, Pratiques* 143-144, 153-167.
- [6] Novak, J.D. 1990. Concept maps and Vee diagrams : Two metacognitive tools for science and mathematics education. *Instructional Science*, 19, 29-52.
- [7] Sullivan, K. P. H. & Lindgren, E. (Eds). 2006. *Computer Keystroke Logging and Writing : Methods and applications*. Amsterdam : Elsevier.