
Écriture et visualisations numériques

Christophe Leblay
christophe.leblay@utu.fi
Université de Turku, Finland

Gilles Caporossi
gilles.caporossi@hec.ca
HEC Montréal, Canada

Depuis les années 80, plusieurs méthodes pour analyser le processus d'écriture ont été utilisées (Miller & Sullivan, 2006). L'outil principal pour analyser le processus d'écriture est le fichier d'enregistrement, appelé *log*, qui contient de façon exhaustive et détaillée l'ensemble des opérations effectuées par le scripteur lors de la rédaction d'un texte (Sullivan & Lindgren, 2014). Les données qui y sont emmagasinées sont considérables, et lorsqu'elles ne sont pas préalablement traitées, elles sont hostiles à l'analyse humaine (Caporossi & Leblay, 2014). Ce traitement préalable nous semble donc déterminant dans l'accès aux données scripturales enregistrées. Autrement dit, il est impossible d'exploiter les programmes d'enregistrement de l'écriture (dits aussi de *temps réel*), sans procéder, au préalable, à un *redéploiement visuel* des données obtenues. Les structures sous-jacentes des données ainsi représentées sont généralement plus propices à l'analyse que les données brutes inexploitable. Plusieurs étapes sont alors nécessaires avant même de pouvoir utiliser les techniques d'analyse à proprement parler. Le processus simplifié par lequel les données sont acquises, traitées et analysées se résume ainsi : a) *enregistrement*, b) *partage*, c) *tri*, d) *recherche*, e) *représentation visuelle des données* et f) *analyse* (Manyika, et al., 2011). Plusieurs de ces actions ont déjà été étudiées de manière individuelle et ont été l'objet de création de logiciels visant à enregistrer le processus d'écriture et à le représenter dans l'une ou l'autre de ses dimensions (Caporossi & Leblay, 2011). Ces logiciels sont tous différents et généralement destinés à un projet de recherche particulier (Sullivan & Lindgren, 2014). Certains sont conçus pour traiter l'information et la retransmettent ensuite à l'utilisateur de façon simplifiée, sous la forme d'une représentation visuelle. L'utilisation de représentations graphiques de données, appelées de façon plus générale *visualisations*, consiste à explorer et essayer de comprendre les grands ensembles de données (Yau, 2011) Elles permettent notamment d'identifier des tendances, structures, irrégularités et relations

entre les données sur une certaine période temporelle (Minelli, et al., 2013, p. 110). Le format compact de la visualisation agrège les données et utilise les capacités cognitives de l'humain (Blanchard, 2005) Le but principal de l'utilisation de ces images représentant le processus scriptural est de laisser l'œil trouver des structures sous-jacentes parmi les données (Tory & Moller, 2004).

Il existe deux grands modèles de représentations qui prétraitent les données de manière à pouvoir appliquer des techniques d'analyse linguistique de données. Le tout premier, attaché aux travaux réalisés par une linguistique cognitive, est celui qui a été développé dans le cadre des *Systèmes d'Information Géographique*, ou SIG ; ont utilisé ce mode de représentation principalement les études suivantes : *LS Graph* (Lindgren & Sullivan, 2002 ; Leijten, et al., 2006), *Genèse du texte* (Doquet-Lacoste, 2003), *GIS Graph* (Lindgren, et al., 2007), *Timeline* (Wengelin, 2009) et *Inputlog* (Leijten, et al., 2013). Le second, celui que nous proposons, en réaction au premier, a été développé en associant, de manière interdisciplinaire, la linguistique génétique et la théorie mathématique des graphes.

La représentation par les graphes (des nœuds reliés entre eux par des liens, ou arcs) permet de mettre davantage en relief l'aspect dynamique de l'écriture (cf. figure 1). Celle-ci est orientée sur la chronologie du processus d'écriture (Leblay, 2011). Un nœud (nommée *cellule* dans notre travail) représente la production d'une suite ininterrompue de frappes au clavier (caractères et espaces).

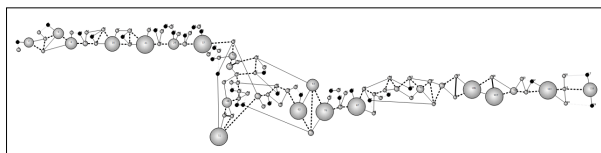


Figure 1. Visualisation globale par graphe : l'écriture exposée (nuance de gris : en sombre = une suppression ; en clair = un ajout)

Ainsi, si l'écriture consistait en une suite ininterrompue de frappes de caractères d'un début jusqu'à une fin, un texte serait visuellement représenté par une seule et unique *cellule* dont la taille dépendrait uniquement du nombre de caractères et d'espaces produits. Or, il existe des retours et des pauses dans ce qui est déjà écrit. Les retours dans le texte sont ainsi marqués: la cellule se divise dès l'instant que la continuité topographique est rompue, bien qu'un lien perdure pour matérialiser le lien temporel (couleurs/nuance de gris et épaisseur du trait). Deux cellules se lient donc quand un lien topographique est créé, le tout au gré des écritures et réécritures.

L'une des particularités de ce travail de visualisation est de bien gérer la transformation et les mouvements du texte grâce aux nœuds représentant des parties de texte (ajoutés, supprimés) reliés ensemble par des liens (ou arcs) définissant leur relation, soit chronologique ou spatiale (Southavilay, Yacef, Reimann, & Calvo, 2013) est possible de voir le contenu de ces nœuds. Telle que mentionnée par Caporossi et Leblay (2011), cette représentation en couleurs, ou en nuances de gris, montre l'aspect temporel de la rédaction, soit le moment exact où le scripteur a effectué chacune des opérations représentées, en lien avec celles qui la précèdent comme avec celles qui la suivent. Les possibilités d'analyse de ces structures ont déjà été étudiées dans le contexte du processus de l'écriture (Leblay & Caporossi, 2014 ; Caporossi & Leblay, 2015).

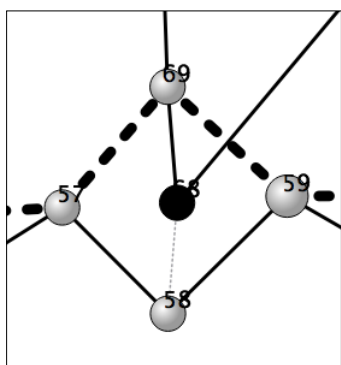


Figure 2. Sous-graphe. Le remplacement

La mise en évidence de sous-graphes particuliers (cf. figure 2) représentant les *patterns* des opérations les plus fréquentes (ajout vs. insertion, suppressions immédiate et différée, remplacement), a déjà été réalisée. L'identification de sous-graphes est utile dans l'analyse globale du graphe représentant le temps de l'écriture (Caporossi & Leblay, 2014)

C'est dans ce cadre que nous proposons 1) un programme dédié, *GenographiX*, pour mettre en évidence tout le travail génétique de réécriture, et 2) un corpus numérique de 10h 20 minutes d'enregistrement ; ce corpus est composé de deux tâches successives, la première de nature narrative (« depuis cette aventure... »), la seconde de nature argumentative (« Qu'est-ce qui est important pour vivre ensemble ? »). Il a été fait le choix de consignes à même de fonctionner auprès de publics très variés (âge, langue maternelle, niveau d'expertise). Précisons que la première reprend un protocole qui a déjà fait l'objet de recueil de corpus de textes auprès d'enfants et de futurs enseignants en formation (Garcia-Debanco et Bonnemaïson, 2014), protocole proposé au départ par Charolles (1988) dans le cadre

d'études de la cohérence textuelle. L'analyse des processus d'écriture enregistrée trouvera ainsi un point d'appui dans ces études consacrées aux produits.

Trois grands axes de recherche émergent alors : (1) le rôle des *opérations génétiques d'écriture-réécriture* qui sous-tendent toute activité scripturale sur papier comme sur écran (Fenoglio, 2012 ; Grésillon, 1994), (2) les différentes chronologies exhaustives qui caractérisent les écritures experte (vs. novice) et universitaire (vs. scolaire), (Bécotte *et al.*, 2016 ; Leblay *et al.*, 2015), et (3) l'impact de la *visualisation* de cette chronologie sur la description et la reproduction de ces écritures dans le cadre de l'enseignement de l'écriture (Doquet & Leblay, 2014 ; Plane *et al.*, 2010). Les résultats que nous proposons, dans cette présentation, concernent particulièrement le phénomène de cohésion textuelle, tel qu'il apparaît, non pas seulement dans un texte produit, mais bien dans son déroulement chronologique. Apparaissent alors des retours ponctuels (qualifiés habituellement d'*erreurs*, d'*écarts*, de *dysfonctionnements*, tout simplement de *variantes*) pris sur le vif, en fonction de différents degrés d'expertise scripturale. Il s'agit de traces génétiques, rendues visibles, de phénomènes de retours dans la construction du sens (cohérence/cohésion) qui n'ont été jusqu'à présent observés et décrits uniquement dans des études basées sur des textes produits.

Ces trois axes nous permettent de souligner la place des visualisations numériques proposées par la génétique textuelle contemporaine dans des dispositifs conçus pour la formation à l'écriture : ces visualisations participent pleinement au travail de réflexivité, à celui de l'articulation entre savoirs théoriques et expérimentations de terrain, et enfin, à celui de la formation de praticiens-chercheurs (Brunel & Rinck, 2017).

Bibliographie

- Bécotte, H.S., Caporossi, G., Hertz, A. & Leblay, C.** (2017). "Writing and rewriting: Keystroke logging's colored numerical visualization." In Sullivan K. P. H. & Lindgren E. (eds), *Observing writing: logging handwriting and computer keystrokes*. Leyde: Brill Academic Publishers. À paraître.
- Blanchard, F.** (2005). *Visualisation et classification de données multidimensionnelles application aux images multicomposantes*. Presses de l'Université de Reims Champagne Ardenne.
- Brunel, M. & Rinck, F.** (2016). "Comment former des enseignants spécialistes de l'écriture et de son enseignement ?" *Pratiques* 171-172, L'écriture professionnelle. <https://pratiques.revues.org/3197>

- Caporossi, G. & Leblay, C.** (2015). "A graph theory approach to online writing data visualization." In Cislaru, G. (ed.) *Writing(s) at the Crossroads: The Process-Product Interface*. Amsterdam: John Benjamins, pp. 171-181.
- Caporossi, G., & Leblay, C.** (2014). "Outils de visualisation de données enregistrées." In Leblay, C. & G. Caporossi, G. (éds.), *Temps de l'écriture: enregistrements et représentations*. Louvain-la-Neuve: Academia, pp. 147-166.
- Caporossi, G., & Leblay, C.** (2011). "Online Writing Data Representation : A Graph Theory Approach." *Lecture Notes in Computer Sciences 7014*, pp. 80-89.
- Charolles, M.** (1988). "La gestion des risques de confusion entre personnages dans une tâche rédactionnelle." *Pratiques* 60, pp. 75-97.
- Doquet-Lacoste, C.** (2003). *Étude Génétique de l'Écriture sur Traitement de Texte d'Élèves de Cours Moyen 2, Année 1995-1996*. Paris: Université Sorbonne nouvelle.
- Doquet, C. & Leblay, C.** (2014). "Temporalité de l'écriture et génétique textuelle: vers un autre métalangage?" In Neveu F. et al. (éds.), *Actes numériques du 4ème Congrès Mondial de Linguistique Française*, CMLF, Berlin, 19-23.07.2014. En ligne : http://www.shs-conferences.org/articles/shsconf/pdf/2014/05/shsconf_cmlf14_01204.pdf
- Fenoglio, I.** 2012 (Ed). *Genesis* 35, Le geste linguistique.
- Garcia-Debanc C. et Bonnemaïson K.** (2014). "La gestion de la cohésion textuelle par des élèves de 11-12 ans : réussites et difficultés." In Neveu F. et al. (éds.), *Actes du Congrès Mondial de Linguistique Française CMLF*, 961-976. En ligne : http://www.shs-conferences.org/articles/shsconf/pdf/2014/05/shsconf_cmlf14_01349.pdf
- Grésillon A.** (1994). *Eléments de critique génétique : lire les manuscrits modernes*. Paris, Presses Universitaires de France.
- Leblay, C.** (2011). *Le Temps de l'Écriture. Genèse, durée, représentations*. En ligne : <https://www.jyu.fi/ajankohtaista/arkisto/2011/11/tiedote-2011-11-04-10-14-59-722468>
- Leblay, C., Caporossi, G., Foucambert, D. & Libersan, L.** (2015). "Écriture & réécriture en situation de travail. Visualisation de pratiques expertes." In Beaudet, C. & Rey, R. (éds.), *Actes du Colloque International L'écriture experte : enjeux sociaux et scientifiques*, Sherbrooke, 13-14.06.2013. Aix-en-Provence: Presses universitaires de Provence, pp. 115-130.
- Leijten, M., & Van Waes, L.** (2006). "Inputlog : New Perspectives on the Logging of On-Line Writing Processes." In Lindgren K. P. (ed.), *Computer Keystroke Logging and Writing*. Elsevier. Pp. 73-94
- Leijten, M., & Van Waes, L.** (2013). "Keystroke Logging in Writing Research : Using Inputlog to Analyze and Visualize Writing Processes." *Written communication* 30(3), pp. 358-392.
- Lindgren, E., & Sullivan, K. P.** (2002). "The LS Graph : A Methodology for Visualizing Writing Revision." *Language Learning*, 52(3), pp. 565-595.
- Lindgren, E., Sullivan, K. P., Lindgren, U., & Spelman Miller, K.** (2007). "GIS for Writing: Applying Geographical Information Systems Techniques to Data Mine Writings' Cognitive Processes." In Rijlaarsdam, G., Galbraith, D., Torrance, M. & Van Waes, L. (eds.), *Writing and Cognition*. Amsterdam: Elsevier. Pp. 83-96
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A.** (2011). *Big data: the next frontier for innovation, competition, and productivity*. McKinsey Global Institute.
- Miller, K. S., & Sullivan, K. P.** (2006). "Keystroke Logging: An introduction." In Sullivan, K. P. & Lindgren, E. (eds.), *Computer keystroke logging and writing*. Oxford: Elsevier. Pp. 1-10.
- Minelli, M., Chambers, M., & Dhiraj, A.** (2013). *Big data, big analytics : Emerging business intelligence and analytic trends for today's businesses*. Indianapolis: Wiley Publishing.
- Plane, S., Alamargot, D. & Lebrave, J.-L.** (2010). "Temporalité de l'écriture et rôle du texte produit dans l'activité rédactionnelle." *Langages* 177, pp. 7-28.
- Southavilay, V., Yacef, K., Reimann, P., & Calvo, R. A.** (2013). "Analysis of Collaborative Writing Processes Using Revision Maps and Probabilistic Topic Models." *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, pp. 38-47.
- Sullivan, K. P., & Lindgren, E.** (2014). "La révision en production écrite enregistrée." In Leblay, C. & Caporossi, G. (éds.), *Temps de l'écriture: enregistrements et représentations*. Louvain-la-Neuve: Academia. Pp. 71-92.

Tory, M., & Moller, T. (2004). "Human factors in visualization research." *IEEE Transactions on visualization and computer graphics*, 10(1), pp. 72-84.

Wengelin, A., Torrance, M., Holmwvist, K., Simpson, S., Galbraith, D., Johansson, V., & Johansson, R. (2009). "Combined eyetracking and keystroke-logging methods for studying cognitive processes in text production." *Behavior Research Methods*, 41(2), pp. 337-351.

Yau, N. (2011). *Visualize this: the flowing data guide to design, visualization and statistics*. Indianapolis: Wiley Publishing.