

Victoria Pérès-Labourdette Lembé

**L'Ecole sous l'influence du numérique :
quelques enjeux et perspectives pédago-techniques**

victoria.pereslabourdette@u-paris10.fr - Laboratoire DICEN-IDF, équipe TACTIC, Université Paris
Ouest Nanterre La Défense

Notre travail interroge ici l'influence du numérique sur l'école, notamment à travers la mise en place de nouveaux lieux de savoirs numériques, les MOOCs. Nous avons mené une expérimentation sur un corpus pédagogique numérique, avec un souci méthodologique identifié comme relevant des Discourse centric learning analytics. Nous montrerons que l'on peut extraire et exploiter des données pédagogiques d'un MOOC sans être partie prenante de celui-ci à un titre autre que celui d'observateur non-participant, et nous nous interrogerons sur les possibilités réflexives d'appropriation des données personnelles induites par cette pratique.

« L'accès à tout pour tous a progressivement quitté son statut de projet idéaliste, conjointement à l'invention d'Internet, et s'est imposé comme un des paradigmes de la société numérique. » (Compiègne, 2010)

Avec Antonio Casilli (2010), nous convenons qu'« il est désormais impossible d'affirmer que les seules vraies communautés sont basées sur un lieu qu'on partage. Les rencontres en face-à-face ne sont plus la modalité exclusive d'interaction. Le contexte sociétal contemporain s'est enrichi de nouvelles manières de « se sentir en communauté » et, par cela, de « faire de la communauté ». Il faut bien souligner ce dernier point : les pratiques communautaires en ligne s'ajoutent à celles qui préexistent et ne se substituent pas à elle ».

L'Ecole, au même titre que l'ensemble des composantes traditionnelles des sociétés contemporaines, évolue sous l'influence du numérique. Les pratiques sociales numériques issues du web 2.0 ont accentué le décalage entre les logiques transmissives et verticales de l'école, et les attentes de reconnaissance, d'implication et de co-construction des apprenants d'aujourd'hui, notamment dans la formation continue.

Ce n'est pas un constat neuf : Martine Fournier et Vincent Troger nous disaient en 2005 que l'ennui général reconnu des apprenants d'aujourd'hui « revêt à présent des formes ostensibles. (...) Il faut apprendre encore à vraiment collaborer (...) Il ne s'agit pas de nier l'individu au profit d'un collectif privilégiant des formats mimétiques. Il n'est donc nullement question d'une intelligence collective comme celle des fourmis, mais de développement individuel au profit du collectif (...), nous ». Mais quelles modalités envisager ?

Le modèle d'un quatrième lieu en ligne, basé sur la théorie du lieu de Ray Oldenburg, questionne les conséquences de l'impact du numérique sur les lieux de savoir traditionnels, qui remet en question la position de l'expert, ainsi que les modalités usuelles de validation curriculaire dans les champs de la connaissance.

Il repose sur :

- l'interrogation des lieux de savoir numériques : les modèles en cours de production, de transmission et de partage des connaissances.
- l'ancrage dans la connexion numérique et la représentation de soi et des structures permises par les TIC.
- l'aide à la constitution et l'accompagnement de communautés numériques d'apprentissage.
- les réflexions sur l'apprentissage rhizomatique de Dave Cormier ; basées sur la théorie du rhizome de Gilles Deleuze et Félix Guattary, elles peuvent y être associées : le modèle pédagogique actuel en formation d'enseignants et de formateurs est très centré sur l'expert, mais la connaissance devient une négociation.

Il convient de porter un regard critique sur ce qui constitue le lieu de la connaissance, et d'interroger les nouvelles façons dont celle-ci peut circuler ou être validée.

En plaçant l'apprentissage social au cœur des lieux de savoir numériques, nous mettons l'accent sur les dispositifs de formation en ligne : le savoir y devient négocié puis validé entre pairs-experts, au travers des communautés numériques d'apprentissage.

Par ailleurs, à l'heure des dispositifs de formation massifiés, le lieu de savoir numérique pose question quant aux perspectives sur l'utilisation de la donnée dans les pratiques pédagogiques contemporaines.

La formation en ligne connaît actuellement d'importantes mutations tant pédagogiques que techniques. Ceci est notamment dû à une massification des usages entraînée par l'industrialisation de la formation en ligne à l'œuvre dans les MOOCs (Massive Online Open Courses).

Ce nouveau mode d'accès au savoir bouscule considérablement les pratiques pédagogiques dans leurs diversités. Source d'innovation pédagogique, il est aussi l'expression de l'influence croissante de l'intelligence artificielle sur les espaces informatisés d'apprentissage humain (EIAH).

Un enseignant faisant face à plusieurs dizaines de milliers d'élèves, ce qui est aujourd'hui une réalité, ne peut envisager d'assurer le suivi pédagogique individualisé de ses apprenants, même accompagné d'assistants (parfois bénévoles).

L'intelligence artificielle accompagne ainsi de plus en plus l'élaboration de dispositifs de formation en ligne massifiés : les traces numériques laissées par les apprenants, au travers des dispositifs numériques qu'ils utilisent, permettent de mesurer, collecter, analyser et communiquer des données sur eux-mêmes et sur leurs contextes, dans l'objectif de mieux comprendre et améliorer les apprentissages et les environnements dans lesquels ils se produisent, et ce à l'aide d'outils informatiques spécialisés permettant de générer des statistiques d'apprentissage (learning analytics en anglais).

L'apprentissage automatique (machine learning en anglais), la discipline scientifique dédiée au développement, à l'analyse et à l'implémentation de méthodes qui permettent à une machine d'évoluer grâce à un processus d'apprentissage, trouve un terrain qui lui est très favorable pour développer la gestion efficace de tâches complexes que l'on pourrait imaginer hors de portée machinique, telles que le tutorat d'apprenant ou la correction de devoirs automatisée (totale ou partielle).

Des récoltes de données d'apprenants sont actuellement moissonnées en volume dans le cadre des MOOCs universitaires pour alimenter l'économie de l'« apprentissage par la donnée » d'aujourd'hui et de demain, le plus souvent essentiellement du fait des prestataires de services auxquelles les structures académiques confient la réalisation de ces formations gratuites. Elles ouvrent l'accès à de nouveaux prospects attirés par ces nouvelles possibilités de qualification (voire de différentes formes de certification, notamment au travers de la badgification), futurs clients de formations diplômantes, pour lesquelles l'investissement requis peut être très important.

Ces données collectées permettent aussi de fournir des modèles d'analyse comportementaux précis et en nombre des apprenants, ce qui accélère les développements scientifiques et techniques de l'apprentissage automatisé.

Ce monde de sciences exactes et d'informatique ne laisse guère de place aux pédagogues, dont beaucoup aujourd'hui ne sont pas informés de ces innovations et mutations en cours.

L'utilisation des données des apprenants à des fins pédagogiques a vocation à être un champ de réflexion pleinement intégré à l'évolution des pratiques pédagogiques contemporaines. Outils et méthodes sont à inventer par les médiateurs du savoir, les spécialistes de l'éducation, et non exclusivement par les spécialistes de l'intelligence artificielle, qui ne sont pas nécessairement formés aux subtilités et aux nécessités de la pratique pédagogique.

Les enseignants et les formateurs peuvent-ils être familiarisés à l'appropriation de ces nouveaux paradigmes par l'expérimentation de pratiques d'analyses pédagogiques quantitatives et qualitatives ?

Nous avons mené une expérimentation sur un corpus pédagogique numérique, avec un souci méthodologique posé comme relevant des Discourse centric learning analytics – mais aussi de la fouille de textes et de l'analyse de discours. Nous avons choisi ce corpus pour sa facilité d'accès, son public visé et son objet pédagogique universitaire, proche de celui de la formation continue académique. La formation continue concerne les personnes qui sont rentrée dans la vie active et/ou qui ont quitté la formation initiale depuis au moins deux ans.

Les Learning analytics constituent une pratique pédagogique récente, évoquée par Philip Long et Georges Siemens en 2011. Les Discourse-centric learning analytics mettent l'accent sur le discours des apprenants pour identifier des modèles d'activité correspondant à un apprentissage et à la construction de la connaissance.

Les MOOCs, actuellement distingués principalement sous deux formes, le xMOOC, académique et transmissif, et le cMOOC, connectiviste, leur fournissent un terrain favorable, et cette pratique se développe en parallèle.

L'enseignement supérieur public français propose ses premiers MOOCs depuis moins d'un an, en particulier au travers de la plateforme de MOOCs France Université Numérique (FUN), et il y a encore peu d'analyses et de travaux de recherche sur ceux-ci, et sur l'utilisation possible et attestée de données fournies par ceux-ci comme matériau d'analyse pédagogique.

Nous montrerons que l'on peut extraire et exploiter des données pédagogiques de ces MOOCs sans être partie prenante de ceux-ci à un titre autre que celui d'observateur non-participant.

Cette captation de données peut se faire dans le respect des règles de la loi française, et en particulier de l'article 2 de la loi Informatique et Libertés, tout en interrogeant les limites de l'anonymat (voulu, subi, rejeté, effectif, illusoire).

La textométrie fournit des outils techniques et méthodologiques qui peuvent être utiles à celui qui veut « faire dire, faire voir » des informations numériques quantitatives ou qualitatives relatives à la donnée discursive issue des espaces de dialogue asynchrones d'un MOOC, ou de tout espace de formation connecté utilisant les fonctionnalités du forum en ligne.

Nous avons analysé un fil de forum sur un MOOC de la plateforme FUN, le MOOC «Quid-Quam ? Eureka ! – Comprendre le monde au quotidien ».

Cette formation en ligne massive et ouverte à tous est proposée par Unisciel.

Unisciel est un Groupement d'Intérêt Scientifique (GIS) regroupant une trentaine d'établissements d'enseignement supérieur français.

Le MOOC «QuidQuam ? Eureka ! – Comprendre le monde au quotidien » a démarré le 12 février 2014 et s'est achevé le 26 avril 2014.

Il est présenté sur la plateforme FUN comme « participatif » et « encourageant l'échange et la discussion autour de 9 thématiques étalées sur 9 semaines ».

Il y est aussi défini comme « ouvert à tous, et (offrant) la possibilité à chacun de poser des questions à laquelle on aimerait avoir la réponse. Une équipe de tuteurs formés, venant des différents établissements, s'occupera d'y répondre au plus vite en collaboration avec les auteurs ».

Le public désigné de ce MOOC est constitué de « toute personne désireuse de satisfaire sa curiosité scientifique. Un enseignant peut également y voir une nouvelle forme d'enseignement originale et complémentaire de son cours présentiel, afin de confronter le programme du cours à la culture scientifique du quotidien. Ce sera le cas notamment à l'Université Bordeaux 1. »

Enfin, la présentation du cours précise qu' « il n'y a pas de prérequis obligatoires pour suivre ce MOOC ».

Selon un communiqué publié sur le site internet de France Université Numérique le 11 mars 2014, il y avait plus de dix mille inscrits à ce MOOC à cette date.

Pour instrumenter nos observations, nous avons utilisé les outils logiciels suivants : Word, Notepad ++, et TXM.

Word est un logiciel de traitement de texte. Notepad ++ est un éditeur de texte.

TXM est un logiciel de textométrie développé par des chercheurs de l'ENS Lyon, de l'Université Lumière Lyon 2 et du CNRS (équipe TXM, Laboratoire ICAR).

La textométrie permet de générer des cartographies synthétiques et visuelles des mots et des textes tels qu'ils s'apparentent ou s'opposent au sein d'un corpus. Elle développe aussi des modèles statistiques pour rendre compte de caractéristiques significatives des données textuelles.

Concernant l'extraction des données, il n'existe pas à ce jour d'API mise à disposition par FUN pour accéder à des sets de données anonymisées. Nous avons donc créé un compte apprenant et procédé à l'extraction manuelle des contenus d'un fil du forum « Déchets et traitement des eaux usées », ouvert par un apprenant et associé à vingt-deux commentaires, en copier / coller dans Word.

L'anonymisation du corpus obtenu a été effectuée manuellement, avec la substitution des éléments suivants par la chaîne de caractères « &lipsum& » :

- Les identifiants des apprenants.
- Les identifiants des membres de l'équipe pédagogique.

- Les identifiants des contributeurs officiels.
- Toute information susceptible de permettre d'identifier une personne physique : ville de résidence, identité civile, ...
- Les mentions de badges des apprenants acquis pendant le cours.
- Les informations temporelles : « environ un mois avant », « -posted environ un mois avant by », « 30 jours avant », « posted 22 jours avant ».
- Les informations pédagogiques : « this post is about PARTIE 5 / Dégradation déchets ».
- Les informations fonctionnelles : « Report Misuse », « Add a comment ».

Il a également été procédé à une suppression :

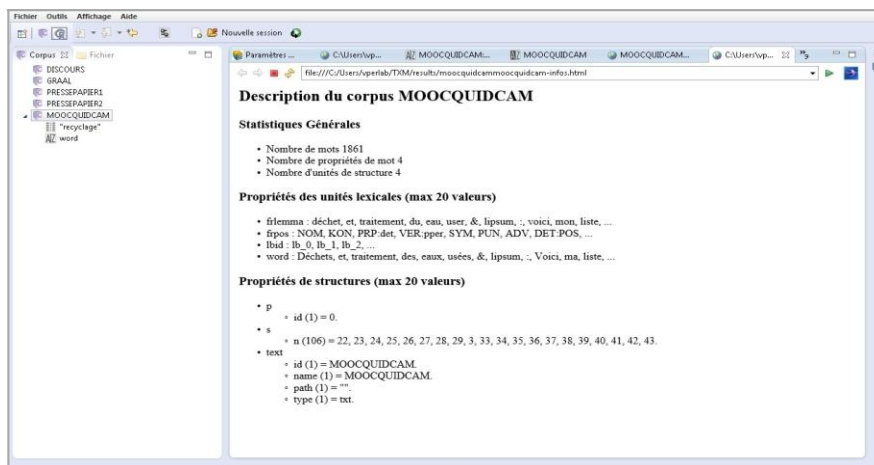
- des liens (le texte lié à un lien est converti en texte brut).
- des images.

Il n'y a pas eu de codage du corpus pour le structurer et le baliser car cela n'était pas utile au traitement envisagé.

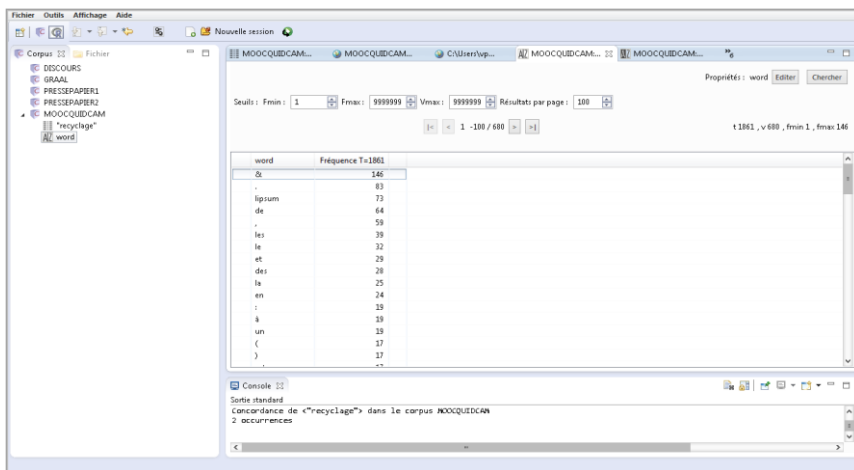
Il n'y pas eu non plus de prise en compte de l'arborescence du forum (sur trois niveaux), ni des attributions de points de vote des utilisateurs du forum pour les commentaires de leurs choix, ni des occurrences d'utilisation de l'icône d'approbation de commentaires apposé sur ceux-ci par l'équipe pédagogique.

Le .doc obtenu a été converti en .txt et importé dans TXM. Le corpus du fil, MOOQUIDCAM, a été créé.

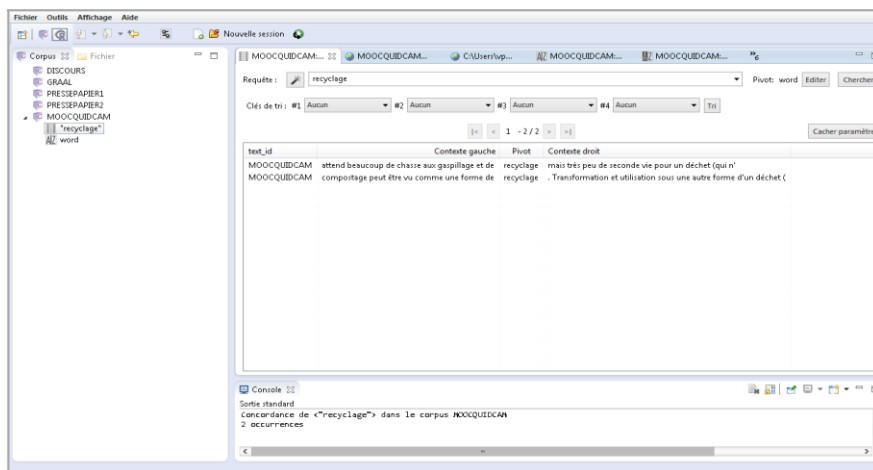
Sa description nous a indiqué qu'il était constitué de 1861 mots.



La commande lexicale a calculé la liste des fréquences de valeurs de propriétés lexicales (ici, la forme du mot) du corpus, ici le fil de forum.



La commande Concordances a cherché les occurrences d'un motif exprimé à l'aide d'une requête CQL et affiche les résultats contextualisés sous forme de concordance KWIC (Key Word In Context).



Les résultats ont été les suivants : hors les mots déterminants, les éléments de la chaîne de caractères d'anonymisation et la ponctuation, on constate que les mots les plus utilisés dans ce fil sont « produits », présents 7 fois, « compostage », présents 5 fois. L'analyse textométrique, selon le site internet de l'équipe TXM, « permet une observation à la fois fine et globale des textes, et donc une exploitation relativement complète des données rassemblées dans ces corpus. De plus, la textométrie reste proche des textes, elle respecte les choix expressifs et les met en valeur : or cette réalité langagière est souvent un terrain d'observation important et très riche pour les sciences humaines et sociales ». L'usage sur un corpus de très petite taille de deux commandes de base de TXM nous montre que la quantification d'éléments textuels produits en ligne en situation pédagogique est possible sans engager l'identification de l'apprenant. Les apports informationnels de cette simple expérimentation sont ici limités, mais les formalisations quantitatives de traces de ce type peuvent selon nous être utilisées pour générer une utilisation réflexive de la part des membres de l'équipe pédagogique comme des apprenants. Ces valeurs, communiquées pendant le MOOC, pourraient accompagner, voire être source, de réflexions / actions pédagogiques, transmissives (tels termes clés inhérents à la notion abordée dans le fil est complètement absent des échanges. On pourra l'aborder plus spécifiquement ultérieurement) ou collaboratives (tels termes clés inhérents à la notion abordée dans le fil est complètement absent des échanges. On pourra penser à l'évoquer aux autres membres de la communauté d'apprentissage à laquelle on participe, par l'écriture sur le forum ou par un autre des canaux de communication mis à disposition dans le MOOC).

Un MOOC est un lieu de savoir numérique qui toujours se veut collaboratif, et qui peut être appréhendé comme un quatrième lieu en ligne.

A l'heure où, comme l'écrit Louise Merzeau, « les données personnalisées doivent trouver leur place dans un nouvel écosystème de services et un nouveau contrat social », des usages pédagogiques reposant sur des analyses discursives quantitatives peuvent selon nous participer à une meilleure appropriation de la trace.

Des études possibles sur ce terrain seraient souhaitables, s'appuyant davantage sur des corpus plus complets et volumineux, et sur l'étendue des possibles fonctions, lectures et visualisations d'un outillage informatique dédié à la fouille de données de plus en plus riche et ouvert, car issu des communautés numériques open source.

Nous achèverons notre exposé en constatant ceci : il existe une tension sociale et pédagogique, issue du paradoxe d'un suivi des apprenants qui occulte volontairement l'identité de ceux-ci, pour mieux se doter de dispositifs de facilitation de l'individuation des apprentissages. Elle doit également être interrogée.

Bibliographie

Casilli A. A., 2010, Les liaisons numériques: vers une nouvelle sociabilité???, Paris, France, Éd. du Seuil, DL, coll.« La Couleur des idées, ISSN 0993-684X », 2010, vol. 1/.

Compiègne I., 2010, La société numérique en question(s), Auxerre, France, Sciences humaines éditions, coll.« La Petite bibliothèque de Sciences humaines, ISSN 1968-8113 », vol. 1/.

Cormier D., 2008, « Rhizomatic Education: Community as Curriculum », Innovate: Journal of Online Education, 4-5..

Deleuze G. et Guattari F., 2002, Mille plateaux, Paris, France, les Ed. de minuit, coll.« Critique (Collection), ISSN 0768-0090 ».

Deuff O. Le, 2011, La formation aux cultures numériques: une nouvelle pédagogie pour une culture de l'information à l'heure du numérique, Limoges, France, Fyp éd., impr. 2011, coll.« Société de la connaissance, ISSN 2259-2857?; [1] », vol. 1/.

Fournier M. et Troger V., 2005, Les mutations de l'école : le regard des sociologues, Auxerre, France, Éditions « Sciences humaines », coll.« Les Dossiers de l'éducation (Auxerre), ISSN 1778-4824 », vol. 1/.

Lafrance J-P, 2009, Critique de la société de l'information, Paris, France, CNRS éd., coll.« Les Essentiels d'Hermès, ISSN 1967-3566 », vol. 1/.

Long P. et Siemens G., 2009, «Penetrating the fog: analytics in learning and education » Educause Review Online 46 (5): 31-40.

Merzeau L., 2013, « L'intelligence des traces », Intellectica, n° 59.

Oldenburg Ray, 1999, *The Great Good Place: Cafes, Coffee Shops, Bookstores, Bars, Hair Salons, and other hangouts at the heart of a community*, Cambridge, MA, Etats-Unis, Da Capo Press, cop.199, vol. 1/.

Pérès-Labourdette Lembé Victoria, 2012, « Bibliothèque quatrième lieu, espace physique et/ou en ligne d'apprentissage social (La) », [http://www.enssib.fr/bibliotheque-numerique/document- 56998](http://www.enssib.fr/bibliotheque-numerique/document-56998)