



HAL
open science

Corpus didactiques : enjeux du traitement des “ traces ” numériques dans l’enseignement distanciel (des langues)

Marc Debono

► To cite this version:

Marc Debono. Corpus didactiques : enjeux du traitement des “ traces ” numériques dans l’enseignement distanciel (des langues). Pierozak Isabelle; Debono Marc; Feussi Valentin; Huver Emmanuelle. Penser les diversités linguistiques et culturelles. Francophonies, formations à distance, migrations, Lambert Lucas, pp.121-134, 2018, 978-2-35935-242-9. hal-02149723

HAL Id: hal-02149723

<https://hal.science/hal-02149723>

Submitted on 6 Jun 2019

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

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

DEBONO, M. (2018), « Corpus didactiques : enjeux du traitement des 'traces' numériques dans l'enseignement distanciel (des langues) », pp. 121-134. PIEROZAK, I., DEBONO, M., FEUSSI, V., HUVER, E. (dir.) (2018), *Penser les diversités linguistiques et culturelles. Francophonies, formations à distance, migrations*, Limoges : Editions Lambert-Lucas, 432 p. URL : <http://www.lambert-lucas.com/livre/penser-les-diversites-linguistiques-et-culturelles-francophonies-formations-a-distance-migrations>

Corpus didactiques : enjeux du traitement des « traces » numériques dans l'enseignement distanciel (des langues)¹

Marc DEBONO
Université François-Rabelais - Tours
EA 4428 DYNADIV

Introduction

La question de plus en plus prégnante de l'utilisation des *corpus didactiques*, en recherche et en formation², emprunte fréquemment des voies technicisantes : ainsi, le modèle du traçage (ou *tracking*) qui se développe surtout – pour le moment - dans le monde anglo-américain (et depuis un certain temps déjà : v. par exemple Hardy, J., Antonioletti, M. et Bates, S., 2004) se donne notamment pour objectif l'amélioration de l'enseignement/apprentissage par une prétendument meilleure analyse des traces des activités des apprenants au sein des dispositifs d'enseignement à distance (EAD désormais).

Il est possible de mettre en perspective cette tendance de l'EAD (et plus particulièrement de l'enseignement distanciel des langues), en la situant au sein de la problématique plus générale du traitement informatique des corpus en SHS, et en sciences du langage (SDL désormais) en particulier (1.). L'outil informatique permettant le traitement technologisé des corpus numériques (qu'ils soient « linguistiques » ou « didactiques » : v. 2. pour une explication de ce qui est entendu par-là) semble renforcer l'évidence d'un sens apparemment plus facilement abordable, accessible et stabilisé. Cette stabilisation du sens des traces numériques (laissées par les étudiants sur une plateforme d'EAD par exemple) permettrait une meilleure prédiction des comportements des acteurs, rendant ainsi possibles des actions (didactiques) plus efficaces³ (3.). Si l'ambition d'objectivation de l'acte didactique dans le but d'en renforcer l'efficacité n'est pas nouvelle en soi et traverse l'histoire de la didactique des langues, la confiance largement partagée en la technologie produit un « effet loupe » sur une tendance à bien des égards problématique : ce texte vise en définitive à explorer ces questions en les portant volontairement sur le terrain épistémologique (4.), et en situant la réflexion, de par les

¹ Ce texte a été repris suite à une communication lors de la journée d'études *Penser en termes qualitatifs le « distanciel formatif »*. En interroger les fondements épistémologiques et didactiques, les visions socio/économico/politiques qui s'est tenue à Tours le 28 mai 2014. Il a bénéficié de la lecture et des commentaires de Véronique Castellotti, Emmanuelle Huver et Didier de Robillard, que je remercie vivement ici.

² A l'heure où nous révisons une ultime fois ce texte, la multiplication des initiatives d'enseignement en ligne des langues – et du FLE en particulier – qui s'est produite me semble confirmer la nécessité d'un débat *sur le fond*, que ces lignes entendent contribuer à faire vivre dans le champ de la DDL.

³ La recherche de la solution la plus efficace en toute chose étant la définition que donne J. Ellul de la technique (v. Ellul, 1954 et 2008).

exemples et références mobilisés, essentiellement dans le champ de l'enseignement distanciel des langues.

1. Les « corpus linguistiques numériques » en SDL

L'équipe de recherche Dynadiv de l'université François-Rabelais de Tours a mené deux réflexions convergentes sur les corpus numériques, dans le cadre de deux projets différents :

- un projet financé par la *Délégation générale à la langue française et aux langues de France* (2013-2014) qui a donné lieu à la parution d'un ouvrage intitulé *Corpus numériques et sens : enjeux épistémologiques et politiques* (Debono, dir., 2014), dans lequel est posée la question du statut des corpus linguistiques numériques en SDL et des enjeux de ce statut.
- le programme *Faire de la Distance un Atout : Constructions Qualitatives du Formatif Distanciel* (2013-2015), axé sur des questionnements didactiques, mais dans lequel la réflexion sur les corpus initiée dans le premier projet trouve un prolongement logique : les questions évoquées à propos des corpus linguistiques numériques recoupant en effet largement celles des corpus didactiques.

Parmi ces questions, une des lignes de force de la réflexion menée par l'équipe Dynadiv sur les corpus linguistiques numériques pourrait être résumée ainsi : l'entrée des SDL dans l'« ère numérique » (avec l'utilisation de plus en plus massive des corpus numériques et de leur traitement automatisé) n'a rien d'une « révolution », mais relève bien plutôt d'une amplification de tendances épistémologiques et d'ambitions déjà bien présentes dans le domaine : empirisme « vérificationniste » (pour reprendre l'expression de C. Taylor, 1997), ambition de prédictibilité et de meilleur contrôle du sens, d'efficacité et d'objectivisme dans l'analyse, corrélé à l'effacement souhaité du chercheur (Debono, 2014 : 13).

Notons que le regain actuel d'un crédit finalement toujours accordé en SDL à ces orientations épistémologiques dépasse largement le champ des SHS et se manifeste par des engouements assez marqués, notamment autour d'une naissante « science des données » (*data science*), qualifiée de « secteur d'emploi le plus sexy du XXI^e siècle » par le très sérieux *New York Times*⁴... Dans ce contexte, la didactique en général et la didactique des langues (DDL désormais) en particulier n'échappent bien entendu pas à cet engouement renouvelé.

2. Les corpus didactiques en DDL : de quelques distinctions

Si l'utilisation des corpus didactiques - c'est-à-dire les corpus compilant des données sur les pratiques des enseignants et/ou des apprenants – a certes toujours existé en DDL, on observe aujourd'hui que l'outil informatique produit un effet loupe sur d'anciennes ambitions de contrôle du « faire didactique ». Autrement dit, l'accentuation actuelle de la technicisation du traitement des « données » didactiques ne veut pas dire que cette velléité ne lui préexistait pas (v. *infra* 3.3.).

Une petite nuance et évolution est tout de même à mentionner, et nécessite de distinguer l'utilisation des corpus de données à des fins de recherche en didactique (ou *didactologie*) et

⁴ Cité par Authemayou, 2014.

l'utilisation des corpus à des fins d'enseignement, avec pour objectif de modifier/améliorer les *pratiques* didactiques⁵.

2.1. En recherche : *didactologie de corpus* et EAD

En didactologie donc (du FLE plus précisément), les corpus didactiques ont toujours été utilisés : on compte ainsi peu de thèses de doctorat en DDL sans constitution de corpus (retranscriptions diverses, d'interactions didactiques principalement, corpus de copies, etc.). Cette utilisation en recherche des corpus à des fins de compréhension des phénomènes d'appropriation des langues (que l'on pourrait qualifier de *didactologie de corpus*, par analogie avec la *linguistique de corpus* à laquelle elle emprunte une orientation empiriste marquée⁶) semble actuellement renforcée par le regain silencieux d'un empirisme épistémologique sous l'effet d'une disponibilité accrue des données « nativement » numériques. Cela est naturellement le cas pour l'EAD : lors du colloque *Données hybrides et contextualisation des corpus. Nouvelles approches* (Paris les 23 et 24 mai 2014), une communication était par exemple intitulée : « *Comprendre l'activité d'apprenants dans un MOOC de français langue étrangère : quelle méthodologie et quels outils pour le recueil et l'analyse de traces en ligne ?* »⁷ (L. Chotel, Université Grenoble 3, laboratoire Lidilem ; je souligne). La « compréhension » (de l'activité des apprenants en ligne) est donc ici, très logiquement dans le paradigme empiriste dominant, étroitement associée aux « traces », recueillies et analysées grâce à des « méthodologies » et « outils ».

Notons que les traces qui intéressent la recherche en didactique des langues sur l'EAD sont par ailleurs très variées quant à leur nature : il ne s'agit pas uniquement de traces discursives (interactions via les forums de discussion, tchats ou mails), mais également extra- et para-linguistiques, comme en témoignent par exemple les travaux sur les corpus d'interactions didactiques via webcam (v. à titre d'exemple : Develotte, Guichon et Vincent, 2010 ; Codreanu et Combe-Celik, 2013) – corpus aisément enregistrables, faciles à « recueillir », « capturables » (v. les utilisations de plus en plus fréquentes des captures d'écrans dans ces travaux), et bientôt peut-être traitables par des logiciels de reconnaissance faciale.

La recherche en DDL mobilise donc des corpus de traces/données depuis ses débuts, mais l'avènement de l'EAD et des « masses » de traces (numériques) qu'il génère, ne fait que renforcer la place de la *didactologie de corpus* dans ce champ disciplinaire.

2.2. Dans les pratiques didactiques : *corpus didactiques* et *corpus linguistiques/langagiers* pour l'enseignement des langues

L'émergence d'une utilisation des corpus de données/traces par les enseignants eux-mêmes (ou plus largement : les praticiens), au niveau des pratiques didactiques donc (et non plus seulement au niveau de la recherche sur ces pratiques) semble plus récente, cette utilisation étant rendue plus aisée avec les NTIC, et en particulier les possibilités de « traçage » qu'elles offrent dans les contextes d'enseignement à distance ou hybrides. Et cette (relativement) nouvelle mobilisation des corpus didactiques numériques vient accompagnée de l'idée - souvent implicite, mais néanmoins agissante - qu'on ne pourra bientôt plus « bien enseigner » les langues sans y avoir recours.

⁵ Il est bien sûr évident que la didactologie vise aussi (même si non exclusivement) à la modification des pratiques didactiques : la distinction a donc ses limites et ne vaut que dans le cadre ici précisé.

⁶ V. Teubert, 2009.

⁷ Programme du colloque disponible en ligne : <http://discdoct.hypotheses.org/558>.

Pour éviter toute possible confusion, il est nécessaire de distinguer l'utilisation des *corpus didactiques* (dont il est question ici) de l'utilisation des *corpus linguistiques/langagiers* pour l'enseignement des langues. Ce qu'on appelle aujourd'hui *data driven learning* (à la suite de Tim Johns, 1991) est l'utilisation des corpus langagiers « authentiques » comme lieu d'exploration par les apprenants, en autonomie : c'est une utilisation didactique des corpus linguistiques/langagiers dans une perspective constructiviste, dans le prolongement des travaux sur l'autonomie, avec à la clé – selon les partisans de cette approche – une rétention améliorée des savoirs linguistiques. En France, les travaux sur cette question au sein du Crapel (Boulton, 2008 ; Landure et Boulton, 2010 ; etc.) prolongent une réflexion ancienne de l'équipe nancéenne sur l'utilisation des documents authentiques en classe et sur l'autonomie (Holec, 1981, entre autres). L'ouvrage *Des documents authentiques aux corpus. Démarches pour l'apprentissage des langues* (Boulton et Tyne, 2014) propose un aperçu récent et complet de la question.

Les utilisations des corpus en DDL qui m'intéresse ici sont d'un autre ordre : il s'agit d'un traçage (en général automatisé par un logiciel) des activités des apprenants constituées en corpus de données (le terme d'origine anglo-américaine aujourd'hui utilisé étant celui de *tracking*⁸) permettant le profilage de ceux-ci en vue d'une intervention didactique efficace : intervention de l'enseignant, de l'apprenant en autonomie, ou de la machine (dans les cas où la réponse didactique est automatisée). Si les apprenants sont aujourd'hui les principaux concernés par ces traçages-profilages, rien n'indique qu'il n'en soit pas bientôt (déjà ?) de même pour les enseignants intervenant sur les plateformes d'EAD (avec, on l'imagine fort bien, une possible facilitation de formes très techniques d'évaluation de leur activité⁹).

3. Corpus didactiques et idéal de prédictibilité des comportements

L'objet de cette partie est d'essayer de mesurer ce que « fait » (implicitement, insidieusement) à la DDL cette utilisation de plus en plus envahissante des corpus didactiques : où cela mène-t-il l'enseignement des langues ? Pour tenter de décrypter les idéaux, imaginaires et croyances qui sous-tendent ce qui pourrait apparaître comme un simple renforcement de l'enseignement par la technique, il n'est pas inutile de partir d'exemples de discours emblématiques et de pratiques d'avant-garde¹⁰, dans le domaine de l'EAD.

3.1. Quelques exemples de discours et de pratiques

Dans le domaine de l'EAD en général...

⁸ La traduction « traçage » ici proposée n'est pas exclusive d'autres traductions possibles, telles que « pistage » ou « traquage ». En revanche le terme de « suivi », traduction également proposée dans certains travaux, me semble très euphémistique, atténuant fortement l'importance accordée aux « traces ».

⁹ Ce qui serait finalement logique au regard du postulat du « traçage », qui est celui d'une efficacité renforcée par le contrôle accru sur les processus d'enseignement et d'apprentissage.

¹⁰ Tout en ayant bien conscience que ces discours, parfois radicaux et exubérants, peuvent facilement servir de « défouloir » à celui qui les conteste, et en masquer d'autres peut-être plus modestes, moins ambitieux, mais mieux implantés. Dans sa réflexion sur la technique, J. Ellul propose une réflexion intéressante : il argumente que le cinéma de science-fiction de son époque (*Alphaville*, 2001, etc.) permettrait à tout un chacun de se « croire conscient » à peu de frais des dangers de l'avancée de la technique, et, par-là, d'en masquer les effets plus modestes, quotidiens. C'est donc en tenant compte de cette réserve que le travail critique qui suit doit être compris.

Daphne Koller est certainement l'une des figures contemporaines les plus en vue de l'enseignement supérieur distanciel : professeure à l'université de Stanford, elle est en effet la fondatrice de *Coursera*, importante plate-forme de MOOCs aux États-Unis. Lors d'une conférence intitulée « Ce que nous apprenons de l'éducation en ligne » (Koller, 2012), elle exposait en ces termes les possibilités qu'offre la collecte des données pour améliorer l'enseignement en ligne : « on peut collecter chaque clic », et de proposer une approche de l'enseignement « massivement » empirico-inductive, dans le but d'en augmenter la fiabilité. Cet idéal de *Coursera* est largement diffusé et explicité :

« La *plate-forme* constitue pour les universités comme *un laboratoire*. Grâce aux données, on peut voir à quel moment du cours *exactement* les élèves ont décroché, quel test ils ont réussi et à quel pourcentage... Dans cette perspective, la technologie a une valeur » (Jake Samuelson, responsable du développement commercial de *Coursera*, cité par Garandau, 2014 ; souligné dans l'article).

Sur la même ligne argumentative, Ben Nelson, fondateur de *Minerva* (projet américain d'université en ligne), affirme que :

« [...] ces cours [en ligne] nous permettront d'accumuler de nombreuses *données sur les comportements et les résultats des étudiants*, dont les professeurs pourront se servir pour ajuster leur enseignement » (cité par Gourdon, 2013 ; souligné dans l'article).

En France, les journées du *Forum Français pour la Formation Ouverte et à Distance*¹¹ (Poitiers, janvier 2014) ont confirmé que la tendance des dispositifs d'EAD¹² était de s'orienter vers de mêmes objectifs de *tracking* que leurs prédécesseurs anglo-américains (selon une logique suiviste peu questionnée).

Malgré l'affichage d'objectifs louables d'amélioration de l'enseignement à distance, tout le monde ne voit pas nécessairement d'un bon œil cet enthousiasme autour du traçage des étudiants par les données, certains s'interrogeant par exemple sur le bien-fondé de l'apparition de nouvelles applications qui permettraient aux professeurs de « savoir » si leurs étudiants avaient lu ou non les textes à lire avant la séance... (Todd Pettigrew, « Technology lets professors track your reading », 2013).

... et dans celui de l'enseignement distanciel des langues : l'exemple du site Frantastique

Le domaine de l'enseignement du FLE n'est pas en reste. En tant que responsable d'un Master FLE, j'ai été contacté en mars 2014 pour la diffusion à nos étudiants d'une offre de stage en création de ressources pour le site *Frantastique*¹³, site d'enseignement du FLE à distance faisant partie d'un groupe plus large d'offre de formation linguistique en distanciel : *Gymglish*¹⁴.

Frantastique propose une formation de FLE en distanciel qui semble avoir peu de concurrents positionnés sur le même créneau pour le moment¹⁵. La formation est assistée par le moteur

¹¹ Association loi de 1901, qui a pour objectif de « permettre aux différents acteurs concernés par la formation ouverte et à distance et le multimédia éducatif » de se rencontrer, réfléchir ensemble, échanger des informations et de collaborer. Source : <http://fffod.fr>.

¹² Tendance à laquelle ne sont pas étrangères les demandes des institutions finançant les dispositifs. Je remercie ici E. Huver et C. Raynal-Astier pour le compte-rendu de ces journées.

¹³ <http://www.frantastique.fr/francais>.

¹⁴ <http://www.gymglish.fr>.

¹⁵ Les rares sites d'enseignement à distance du FLE existants sont généralement à but non lucratif et sensiblement différents dans leurs conception ou objectifs : on pourrait par exemple citer « Le français en

d'intelligence artificielle *A9expert*, « qui élabore et personnalise les leçons envoyées chaque matin à chaque utilisateur », moteur alimenté par les traces laissées par les étudiants sur la plateforme (« *A9expert* détecte les lacunes et besoins au travers des réponses de chaque utilisateur. Il compose de façon quasi-immédiate une correction personnalisée et commentée avec le score du jour »¹⁶). L'idéal du traçage des étudiants se retrouve également dans le pack proposé aux universités qui voudraient utiliser ce programme, avec un outil de « supervision » (de suivi) - en développement - qui permettrait de suivre précisément l'activité de chaque étudiant ou groupe d'étudiants sur un ensemble précis de points :

« Supervision tools and a dedicated web space are provided for the teachers and administrators. This space provides a snapshot of students' progress within the French course, and includes detailed information about scores, participation rates, and more, *allowing teachers to track and assess each student's progress easily, individually and within a group*. Students are evaluated continuously throughout the lessons on a scale of 0 (beginner) to 5 (professional mastery of international French) » (je souligne)¹⁷.

Ce site d'enseignement du français en ligne met donc l'accent sur le suivi (argument essentiel [aujourd'hui](#) dans le domaine du distanciel), mais suivi d'un genre particulier : celui-ci étant d'abord garanti par l'activité d'une machine intelligente traitant des données (le moteur d'intelligence artificielle *A9expert* et les « outils de supervision »), et, dans un second temps uniquement, par une équipe d'enseignants. La base de toute intervention didactique est donc constituée par l'ensemble des traces laissées par les apprenants travaillant sur la plateforme, préalablement traitées par la machine.

3.2. Les *learning analytics* : idéal de prédictibilité et données

Ces discours et pratiques entourant des projets d'EAD d'envergures diverses (*Coursera*, *Minerva*, *Frantastique*, etc.) peuvent utilement être rapprochés de ce qui est actuellement travaillé en didactique (en sciences de l'éducation plus particulièrement) sous l'appellation de *learning analytics* (*LA* désormais), dont Brown donne la définition suivante :

« Analytics is an umbrella term for the use of data, statistical analysis, and explanatory and predictive models to gain insights and act on complex issues. [...]. [*LA* are defined as] « the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs » (Brown, 2012 : 1).

L'ambition finale des *LA* est de permettre une meilleure gestion du travail étudiant en produisant des réponses pédagogiques « automatiques » et « semi-automatiques », à partir de

première ligne », plateforme développée par l'Université de Grenoble et qui s'inscrit dans le cadre d'un dispositif de recherche-action et de formation de futurs enseignants, dans le cadre de Masters FLE (<http://fle-1-ligne.u-grenoble3.fr>) ; ou encore « Le français en ligne » développé par l'Institut français de Grèce, qui s'apparente plus à un ENT (<http://www.lefrancaisenligne.gr>). Quant au premier MOOC de FLE apparu en 2014, « Travailler en français », il se distingue par son modèle connectiviste, son caractère très ciblé (« aider à acquérir des connaissances et des savoir-faire en français pour rechercher un emploi en France ou dans un pays francophone ») et limité dans le temps (<https://sites.google.com/site/mooctravaillerenfrancais/home>).

¹⁶ <http://www.frantastique.fr/francais>.

¹⁷ <http://www.frantastique.com/learn-french-offers/universities>. Ces présupposés didactiques sont traduits de manière moins précise sur la version française du site : « Des outils de supervision sont mis à la disposition des professeurs et des administrateurs. Cet espace donne accès à des informations concernant le suivi de la formation, l'assiduité, les progrès réalisés... Les étudiants sont évalués en continu tout au long des leçons sur une échelle de 0 (débutant) à 5 (maîtrise professionnelle du français international) » : <http://www.frantastique.fr/apprendre-francais-offre/universite>.

l'analyse des traces que l'étudiant aura laissées, ceci dans le but (louable) d'orienter son travail et d'éviter son « décrochage ». Parmi les réponses automatiques, on trouve un panel assez large : des simples « alertes » par codes de couleurs adressées à l'apprenant au tutorat assisté par des systèmes d'intelligence artificielle. Les réponses semi-automatiques sont celles qui passent d'abord par un signal envoyé (par la machine) à l'équipe pédagogique, laquelle décide du type de réponse à apporter dans un second temps. C'est donc toujours l'analyse automatisée du corpus didactique par une machine intelligente qui déclenche l'intervention didactique, quelle que soit sa nature (v. Brown, 2012 : 4-5).

Il est assez remarquable de constater que les *LA* insistent sur la séparation entre ce qui relèverait de la *technologie* (l'analyse des données, des corpus numériques), et ce qui relèverait des *objectifs*, c'est-à-dire la « pédagogie », de « l'humain » : les technologies d'analyse des données *servent* la pédagogie, mais le plus important vient après, l'essentiel étant dans l'action consécutive à ces analyses fortement technologisées. L'analogie est ainsi faite entre l'intervention de l'enseignant et celle du médecin : le « portrait » de l'apprenant donné par l'analyse des données didactiques serait comparable à celui du patient donné par les analyses médicales¹⁸. Il y a là un cas typique de réduction de la technologie à un simple outil (qu'il faudrait, évidemment, « bien » utiliser) qui serait au service de l'enseignant, lequel resterait en définitive le dernier décideur puisqu'il en a la maîtrise. Cette réduction masque pourtant les effets plus profonds de cette technologie du traçage/*tracking* sur la pédagogie, sur la conception de l'apprenant et de l'enseignant, et par extension sur la conception de l'humain. L'ambition de rendre prédictibles les étudiants (et pourquoi pas, bientôt, les enseignants) va en effet tellement loin que l'on peut légitimement s'interroger sur la conception de l'humain et de sa diversité qui la sous-tend : nous avons bien affaire ici à *une* conception de l'humain comme idéalement prédictible, a minima questionnable (v. *infra* 4.)¹⁹.

En définitive, nous ne pouvons que constater que cette ambition de prédictibilité des comportements étudiants (« *predictive models* ») portée par les *LA* et appuyée sur les moyens techniques de traitement quantitatif, statistique et largement automatisé des données, est très caractéristique des tendances de l'EAD (des langues notamment) exemplifiées *supra* (3.1.). Néanmoins, il ne faudrait pas non plus donner l'illusion que ces tendances apparaissent avec l'avènement des nouvelles technologies de l'information et de la communication (NTIC) pour l'enseignement des langues : d'abord parce que ces modèles sont réminiscent de leurs origines intellectuelles (behaviorisme, cognitivisme), ensuite parce que la DDL n'en est pas à son coup d'essai dans sa quête de « l'apprenant prédictible ».

3.3. Continuités en DDL et « effet loupe »

¹⁸ « As we have seen, the work reported by Essa and Ayad incorporates « a case-based methodology for managing interventions », *analogous to the way cases are managed in the medical domain. It provides an up-to-date history of a learner's interactions with faculty, advisors, and other contacts in student services, all serving to deliver a more detailed and nuanced « portrait » of the learner* » (Brown, 2012 : 4 ; je souligne). Brown renvoie ici à ce travail : Essa, A. et Ayad, H., 2012, « Student Success System: Risk Analytics and Data Visualization using Ensembles of Predictive Models », paper presented at the Learning Analytics and Knowledge Conference, Vancouver, British Columbia, April 29–May 2, 2012.

¹⁹ Nous ne pouvons également ici que renvoyer aux nombreux travaux dans la littérature francophone actuelle qui traitent de cette question de la (re-)définition de « l'humain » en lien avec les nouvelles technologies (entre autres, la liste étant loin d'être exhaustive : Breton, 1992 ; Lafontaine, 2004 ; Besnier, 2012). Dans le champ de l'EAD, nous (avec mes collègues Isabelle Pierozak et Didier de Robillard) avons présenté au colloque *Epal 2013 (Échanger pour apprendre en ligne, Grenoble)* une communication intitulée « Quelle technicité pour quelle humanité en FOAD ? Premiers éléments de réflexion épistémologique en lien avec une recherche-intervention » qui entendait interroger les conceptions de l'humain sous-jacentes aux initiatives d'enseignement en ligne (des langues notamment).

En effet, à bien y réfléchir, cet idéal de prédictibilité des comportements est certes renforcé par les multiples croyances actuelles dans les possibilités qu'offre la technologie (c'est en ce sens que l'on peut parler d'un « effet loupe »), mais n'est finalement que la continuation de tendances déjà présentes dans l'histoire de la DDL.

Pour ne prendre, à titre d'exemple, que quelques jalons bien connus de l'histoire de la DDL : l'inspiration béhavioriste de la *MAO* (méthode audio-orale) américaine qui vise à conditionner le comportement de l'apprenant ; le modèle structural de la méthode *SGAV* (structuro-globale audiovisuelle) française qui cherche à travailler l'appropriation linguistique à partir de la prédictibilité structurale des comportements langagiers ; dans la plus récente *approche communicative*, l'acte de langage est un acte maîtrisé quant à son intentionnalité et ses effets, entrant dans une communication dont les éléments sont idéalement prédictibles²⁰ ; la *perspective actionnelle* qui lui a succédé s'appuyant de ce point de vue sur des options similaires, en les élargissant à l'action « sociale » (et non plus seulement « langagière ») ; enfin, l'influence grandissante des sciences cognitives en DDL – que l'on trouve par exemple dans des développements comme l'*ANL* (approche neuro-linguistique : v. Netten et Germain, 2012 ; Germain, 2017) ou la « didactique cognitive des langues » (Billères et Spanghero-Gaillard, 2005) – s'appuie sur l'idée d'une possible prédiction des comportements d'appropriation linguistique par l'étude du fonctionnement cérébral ou de la « boîte noire »²¹ de la pensée²².

D'une certaine manière, l'utilisation des traces/données dans l'enseignement distanciel des langues ne fait donc qu'extrapoler des ambitions de prédiction comportementale inscrites dans l'histoire même de la DDL (et dans la constitution corolaire d'un champ de recherche scientifique : pour « être science », dans le paradigme dominant, il faut pouvoir *prédire* : v. Taylor, 1997). Aussi, pour essayer de comprendre (et de contrebalancer) l'engouement actuel, à l'heure des *big data*, pour les corpus numériques en DDL, il est nécessaire d'en passer par une mise en question, sur le plan épistémologique, d'un idéal techniciste qui traverse la discipline didactique, et qui oppose fortement la prédictibilité au flou de « l'intuitif » – j'emprunte ici à dessein le terme utilisé par G. Vigner (2012), qu'il met en regard de l'évolution, selon lui déplorable, de la formation des enseignants de langues vers toujours plus de technicisation.

²⁰ Sur les liens entre approche communicative et pragmatique linguistique, voir Debono, 2013.

²¹ Cette métaphore de la « boîte noire » est empruntée à un autre texte sur la question, en didactique du français langue maternelle : « Didactique du français langue maternelle : approche(s) "cognitivist(e)s" ? » (Brassart, 2004 : 97-98).

²² On peut mentionner à cet endroit la pertinente mise en garde de D. Simard contre l'influence du cognitivisme en didactique, qui rejoint à certains égards la critique menée ici : « [...] qu'il s'agisse des sciences de l'éducation ou des milieux scolaires, le cognitivisme a le vent en poupe [...]. A l'instar du béhaviorisme qu'il prétend remplacer, le cognitivisme propose une vision scientifique et technologique de l'éducation. *L'enseignement et l'apprentissage sont des activités de traitement de l'information* ; elles reposent sur des processus cognitifs qui sont assimilables au fonctionnement computationnel d'un ordinateur. *Si, pour le behaviorisme l'être humain est un pigeon compliqué, pour le cognitivisme, l'esprit humain est un ordinateur compliqué.* Bien que je reconnaisse la fécondité du cognitivisme et ses retombées positives sur la formation et la pratique enseignantes, je m'inquiète toutefois du réductionnisme où il risque de tomber. Il faut donc rappeler avec Bruner (1991) que *l'esprit prend forme au travers de l'histoire et de la culture* et que pour mieux comprendre l'esprit humain, son activité créatrice et ses produits culturels, il nous faut réintroduire dans l'éducation une *approche historique, culturelle et interprétative*. Depuis Heidegger et Gadamer, l'herméneutique comprend l'existence humaine dans son rapport au monde comme interprétation, c'est-à-dire comme expérience que se réalise sur le mode d'un échange dialogique au sein d'une langue. Tout homme interprète le monde et lui donne une signification à partir d'une histoire et d'une culture où il s'inscrit comme être humain » (Simard, 2004 : 120 ; je souligne).

4. En guise de conclusion : une indispensable critique de l'idéal techniciste du traçage/traque par les données

4.1. Éviter l'écueil de la moralisation du débat

La plupart des dispositifs de traçage mis en œuvre dans l'EAD ne sont pas conçus dans une perspective « surveiller et punir » (crainte majeure quant aux possibles dérives liberticides de l'utilisation des données numériques et autres *big data*), mais bien plutôt surveiller pour mieux accompagner, pour éviter le décrochage des apprenants, etc. Autant de visées louables, allant dans le « sens du progrès²³ » et de l'efficacité accrue (le site *Frantastique* se vante par exemple de ces résultats en « rétention »), rendent une critique de ces dispositifs difficilement audible (si c'est pour le bien de l'apprenant...) et l'accusation de technophobie est en général de mise. Par ailleurs, si la critique reste sur un plan éthique ou moralisant (il ne faut pas réduire l'humain aux traces²⁴, il ne faut pas « fliquer » les apprenants en distanciel, etc.), elle risque de tourner rapidement court, tant l'unanimité sur ces questions est de bon ton.

S'attacher à mettre en évidence l'idéal de prédictibilité – et donc la vision de l'humain – qui fonde l'utilisation corpus didactiques dans l'EAD ne relève donc pas tant du plan moral/moraliste (prétentions louables vs liberticides) que d'une interrogation sur les présupposés épistémologiques sur lesquels reposent de tels dispositifs : s'intéresser aux « racines » plutôt qu'aux « feuilles » donc.

4.2. S'intéresser aux « racines » plutôt qu'aux « feuilles », en portant le débat sur un terrain épistémologique

Aborder ainsi les choses permet de dépasser les stériles oppositions moralisantes entre technophobes et technophiles. L'utilisation ou non des technologies n'est pas le problème : nous les utilisons, mais en refusant qu'elles « nous changent » (Ellul, 1954), qu'elles « nous arraisonnent » (Heidegger, 1953, 1962)²⁵. Cela suppose un travail de théorisation important que nous avons essayé de mener pour les « corpus numériques linguistiques », en ancrant nos réflexions dans la philosophie herméneutique et la phénoménologie (v. les contributions de L. Bonoli, M. Debono, V. Feussi, I. Pierozak, E. Razafi et D. de Robillard, réunies dans Debono, dir., 2014). Ces courants (dont Heidegger, Gadamer, Ricœur, Taylor, etc. sont les représentants les plus connus) constituent en effet de puissantes alternatives pour remettre en cause la perspective d'une « traque » de « chaque clic » (en didactique du distanciel), ou de chaque interaction sur les réseaux sociaux (en analyse conversationnelle par exemple). En affirmant que tout le sens possible n'est pas dans

²³ Si l'on considère le « progrès » dans le sens particulier de *puissance de prévision*, exposé ici, dans une perspective critique, par Valéry : « Je me suis essayé autrefois à me faire une idée positive de ce que l'on nomme progrès. Éliminant donc toute considération d'ordre moral, politique ou esthétique, le progrès me parut se réduire à l'accroissement très rapide et très sensible de la puissance (mécanique) utilisable par les hommes, et à celui de la précision qu'ils peuvent atteindre dans leurs prévisions » (Paul Valéry, « Propos sur le progrès », 1929, cité par Grinevald, J., « Progrès et entropie, cinquante ans après », in Bourg, D. et Besnier, J.-M., (dir.), *Peut-on encore croire au progrès ?*, Paris : PUF, 2000, p. 220).

²⁴ Sur ce point, on trouvera plusieurs contributions intéressantes dans le n°68 de la revue *Hermès* : « L'Autre n'est pas une donnée » (2014). Voir également : Galinon-Méléneq, B., 2012, « Le numérique. De l'usage aux traces. Du fantasma au doute et au cauchemar », in Denoit, N., *L'imaginaire et la représentation des NTIC dans la société contemporaine*, PUF.

²⁵ Si les deux réflexions sur la technique (Ellul, Heidegger) sont bien sûr fort différentes quant à leurs arrière-plans respectifs – qu'il serait trop long d'exposer ici – elles partagent cette idée d'une non-neutralité de la technique moderne.

les traces visibles, que même si l'on « voyait » tout on ne comprendrait pas tout (l'humain n'étant totalement transparent ni à lui-même, ni aux autres), l'herméneutique et la phénoménologie nous aident à aller à l'encontre de l'idée que la seule manière de comprendre (l'appropriation d'une langue par un apprenant sur une plateforme d'enseignement à distance par exemple) est d'en passer par la collecte et l'analyse exhaustive et multimodale de « données brutes », alors considérées comme « données de base de la connaissance » (Taylor, 1997 : 144).

Les orientations essentiellement empirico-inductives des SHS ont largement contribué à l'importance disproportionnée accordée aux traces/données/corpus dans la compréhension des comportements humains. Contrebalancer cette surenchère empiriste (qui se manifeste actuellement dans l'EAD, en DDL et en SDL pour ne mentionner que nos domaines d'intervention) nécessite de favoriser la diversité des interprétations que présuppose la non-transparence de l'être humain. Cela passe par une nécessaire mise en question de l'importance des signes/données/traces dans la constitution du sens – ce à quoi les SHS empiristes et « vérificationnistes » sont par définition rétives (v. Taylor, 1997) – et la revalorisation de l'expérience « antéprédicative », non-discursive et non exclusivement rationnelle, explicite, stable, ni homogène (v. Romano, 2010) : cette expérience est, pour les courants phénoménologiques et herméneutiques, première et fondamentale en ce qu'elle *rend possible* la raison sémiotique, qui, elle, est seconde (sur ce point, voir : Robillard, 2012, 2014)²⁶.

Plus qu'un « droit à l'opacité » (avancé dans le débat actuel sur le *big data*), il s'agit d'affirmer, en la fondant épistémologiquement dans les courants phénoménologiques et herméneutiques, la nécessaire reconnaissance d'une intériorité constitutive et inaccessible de l'être humain, intériorité métaphorique servant à signifier que des choses « échappent » nécessairement à la compréhension. Or c'est exactement cela que l'utilisation des corpus didactiques tend à effacer insidieusement, sans le dire, sous couvert de technicisation des pratiques. De ce point de vue, le problème posé ici à propos de l'EAD renvoie à un débat de société qui dépasse très largement les seules questions d'enseignement et d'apprentissage.

Néanmoins, modestement mais sûrement, il y a beaucoup à faire du côté de la formation des enseignants de langues. S'il est en général assez aisé de faire comprendre les dérives d'une excessive logique de *tracking* avec des arguments moralisants, en discuter les arrière-plans épistémologiques est plus difficile : c'est néanmoins ici que l'essentiel se joue, et cette conviction préside à l'orientation de nos formations de master en didactique du FLE/S à l'université François-Rabelais de Tours. Sans être totalement fataliste, il est pourtant probable que la logique du traçage/profilage des apprenants (et enseignants, bientôt) par les données a un bel avenir devant elle – dans la mesure où elle s'inscrit dans une continuité historique (recherche de prédictibilité et d'assise « scientifique ») –, en particulier dans ce contexte de l'EAD où il n'y a qu'à se baisser pour récolter des traces...

Éléments de bibliographie

²⁶ « Cette démarche est beaucoup plus à contre-courant et choquante pour la tradition de la pensée scientifique que celles que les adversaires du scientisme étroit adoptent généralement » (Taylor, 1997 : 152), et il ne faudrait pas la confondre avec certaines mobilisations de l'herméneutique qui restent centrées sur les signes-données-traces : en SDL, on peut mentionner « l'herméneutique matérielle » de F. Rastier (2005), « l'herméneutique numérique » de D. Mayaffre (2002), ou encore, la linguistique de corpus fondée sur l'« herméneutique » (philosophique) proposée par W. Teubert (2009).

Albero Brigitte et Thibault Françoise, 2009, « La recherche française en sciences humaines et sociales sur les technologies en éducation », *Revue française de Pédagogie*, n° 169, p. 53-66.

Authemayou Céline, 2014, « Le big data, une masse de données qui vaut de l'or », publié le 7 février 2014 sur *EducPros.fr*.

Besnier Jean-Michel, 2012, *L'Homme simplifié : Le syndrome de la touche étoile*, Paris, Fayard.

Billières Michel et Spanghero-Gaillard Nathalie, 2005, « La didactique cognitive des langues : regards croisés de disciplines pour comprendre le "comment" », *Parole*, n°34-36, p. 101-136.

Boulton Alex, 2008, « Esprit de corpus : promouvoir l'exploitation de corpus en apprentissage des langues », *Texte et corpus*, n° 3, p. 37- 46.

Boulton Alex et Tyne Henry, 2014, *Des documents authentiques aux corpus. Démarches pour l'apprentissage des langues*, Paris, Didier, coll. « Langues et didactique ».

Brassart Dominique-Guy, 2004, « Didactique du français langue maternelle : approche(s) "cognitivist(e)s" ? », dans J.-L. Chiss, J. David et Y. Reuter, *La didactique du français*, Bruxelles, De Boeck, p. 95-118.

Breton Philippe, 1992, *L'utopie de la communication. Le mythe du 'village planétaire'*, Paris, La Découverte, 2011.

Brown Malcolm 2012, « Learning Analytics : Moving from Concept to Practice », *EDUCAUSE Learning Initiative* (<http://net.educause.edu/ir/library/pdf/ELIB1203.pdf>).

Bruillard Eric, 1997, « L'ordinateur à l'école : de l'outil à l'instrument », dans L.-O. Pochon et A. Blanchet (éds.), *L'ordinateur à l'école : de l'introduction à l'intégration*, Neuchâtel, IRDP : LEP, p. 101-118.

Codreanu Tatiana et Combe-Celik Christelle, 2013, « Effects of webcams on multimodal interactive learning », *ReCALL*, vol. 25, n°1, p. 30-47.

Debono Marc, 2013, « Pragmatique, théorie des actes de langages et didactique des langues-cultures. Histoire, arrière-plans philosophiques, conséquences et alternatives », dans V. Castellotti (sous la dir. de), *Le(s) français dans la mondialisation*, Fernelmont, Éditions Modulaires Européennes, p. 423-447.

Debono Marc (dir.), 2014, *Corpus numériques, langues et sens : enjeux épistémologiques et politiques*, Bern, Peter Lang, coll. Gram-R – Études de linguistique française (dirigée par V. Van Raemdonck), n°25.

Debono Marc, 2014, « Introduction. Retours sur un appel à proposition. 'Pratiques' et 'représentations' en sciences du langage », dans M. Debono (dir.), *Corpus numériques, langues et sens : enjeux épistémologiques et politiques*, Bern, Peter Lang, coll. Gram-R – Études de linguistique française (dirigée par V. Van Raemdonck), n°25, p. 11-22.

Develotte Christine, Guichon Nicolas et Vincent Caroline, 2010, « The use of the webcam for teaching a foreign language in a desktop videoconferencing environment », *ReCALL*, vol. 23, n°3, p. 293–312.

Ellul Jacques, 1954 [1990], *La Technique ou l'Enjeu du siècle*, Paris, Economica.

Ellul Jacques, 2008, *Ellul par lui-même. Entretiens avec Willem H. Vanderburg*, Paris, La table ronde (entretiens réalisés en 1979).

Endrizzi Laure, 2012, « Les technologies numériques dans l'enseignement supérieur, entre défis et opportunités », *Dossier d'actualité Veille et Analyses*, n°78 (<http://ife.ens-lyon.fr/vst/DA/detailsDossier.php?parent=accueil&dossier=78&lang=fr>).

Galinon-Méléneq Béatrice, 2012, « Le numérique. De l'usage aux traces. Du fantasme au doute et au cauchemar », dans Nicole Denoit, *L'imaginaire et la représentation des NTIC dans la société contemporaine*, Tours, PUFR, p. 79-100.

Galinon-Méléneq Béatrice et Zlitni Sami (dirs), 2013, *Traces numériques*, Paris, CNRS Editions.

Garandeau Coralie, 2014, « Off to Silicon Valley (3/10). Coursera veut réfléchir à l'université de demain », *EducPros.fr* (<http://www.letudiant.fr/educpros/entretiens/off-to-silicon-valley-3-10-coursera-veut-reflechir-a-l-universite-de-demain.html>).

Germain Claude, 2017, *L'approche neurolinguistique*, Longueuil (Canada), Myosotis Presse.

Gourdon Jessica, 2013, « États-Unis : Minerva veut bousculer les frontières de l'université "on-line" », *L'étudiant* (<http://www.letudiant.fr/educpros/entretiens/etats-unis-minerva-veut-bousculer-les-frontieres-de-l-universite-on-line.htm>).

Guichon Nicolas, 2006, *Langues et TICE : Méthodologie de conception multimédia*, Paris, Ophrys.

Hardy Judy, Antonioletti Mario et Bates Simon, 2004, « e-Learner tracking : tools for discovering learner behaviour », *The IASTED International Conference on Web-Based Education*, Austria, p. 458-463.

Holec Henti, 1981, *Autonomy and Foreign Language Learning*, Oxford: Pergamon.

Heidegger Martin, 1953 [1973], « La question de la technique », dans *Essais et conférences*, Paris, Gallimard, p.9-48.

Heidegger Martin, 1962 [1990], *Langue de tradition et langue technique*, Hamburg, Lebeer-Hossmann.

Johns Tim, 1991. « Should you be persuaded : two samples of data-driven learning », in T. Johns, T. et P. King (eds), *Classroom Concordancing : English Language Research Journal*, n°4, p. 1-16.

Koller Daphne, « Ce que nous apprenons de l'éducation en ligne », conférence TED de 2012, *L'étudiant* (<http://www.letudiant.fr/educpros/actualite/video-reinventons-l-ecole-les-meilleures-conferences-ted-sur-l-innovation-pedagogique.html>).

Lafontaine Céline, 2004, *L'empire cybernétique. Des machines à penser à la pensée machine*, Paris, Seuil.

Landure Corinne et Alex Boulton, 2010, « Corpus et autocorrection pour l'apprentissage des langues », *ASp. La revue du GERAS*, n°57, p. 11-30.

Linard Monique, 1996, *Des machines et des hommes : apprendre avec les nouvelles technologies*, Paris, L'Harmattan.

Mayaffre Damon, 2002, « L'Herméneutique numérique », *L'Astrolabe*, n° spécial (http://halshs.archives-ouvertes.fr/hal-00586512_v1), p. 1-11.

Netten Joan et Germain Claude, 2012, « Un nouveau paradigme pour l'apprentissage d'une langue seconde ou étrangère : l'approche neurolinguistique » (http://www.ambafrance-cn.org/IMG/pdf/netten_et_germain_-_un_nouveau_paradigme_-_anl.pdf), Traduction de Netten Joan et Germain Claude, 2012, « A new paradigm for the learning of a second or foreign language: The neurolinguistic approach », *Neuroeducation*, 1(1) (<http://www.neuroeducationquebec.org/revue>), p. 85-114.

Paquien-séguy Françoise, 2006, « Les technologies d'information et de la communication : constat, questionnements et hypothèses », HAL (https://hal.archives-ouvertes.fr/sic_00104288/document)

Pettigrew Todd, 2013, « Technology lets professors track your reading », (<http://oncampus.macleans.ca/education/2013/04/15/technology-lets-professors-track-your-reading>)

Pierozak Isabelle (éd.), 2010, *Du « terrain » à la relation : expériences de l'internet et questionnements méthodologiques*, *Cahiers de linguistique - Revue de sociolinguistique et de sociologie de la langue française*, n°36/2.

Robillard (de) Didier, 2012, « Diversité, sens : enjeux », dans M. Debono et C. Goï (dirs.), *Regards interdisciplinaires sur l'épistémologie du divers. Interculturel, herméneutique et interventions didactiques*, Fernelmont, Éditions Modulaires Européennes, p. 195-216.

Robillard (de) Didier, 2014, « Monnaie de signe, monnaie de singe ? Comment comprendre des corpus électroniques? Implications épistémologiques, éthiques et politiques », dans M. Debono (dir.), 2014, *Corpus numériques, langues et sens : enjeux épistémologiques et politiques*, Bern, Peter Lang, coll. Gram-R – Études de linguistique française (dirigée par V. Van Raemdonck), n°25, p. 145-214.

Romano Claude, 2010, *Au cœur de la raison, la phénoménologie*, Paris, Gallimard.

Simard Denis, 2004, *Éducation et herméneutique : contribution à une pédagogie de la culture*, Laval, Presses de l'Université de Laval.

Taylor Charles, 1997, *La liberté des modernes*, Paris, PUF, Chapitre : « L'interprétation et les sciences de l'homme », p.137-194.

Teubert Wolfgang, 2009, « La linguistique de corpus : une alternative », *Semen* [En ligne], 27 | 2009, (<http://semen.revues.org/8914>)

Vigner Gérard, 2012, « Formation des enseignants et histoire de la diffusion et de l'enseignement des langues », *Le français dans le monde. Recherches et applications*, n°52, juillet 2012, p. 78-92.