



ISSN 2260-1651

ISSN en ligne 2260-4987

## Corpus web 2.0 : quelques enjeux méthodologiques et épistémologiques

**Sabrina Bevilacqua**

Universidad de Buenos Aires (UBA) - IESLV Juan R. Fernández,  
Argentine  
sabinabevilacqua@yahoo.com.ar

Reçu le 15-10-2015 / Évalué le 11-11-2015/ Accepté le 18-04-2016

### Résumé

Nombre de défis d'ordre méthodologique et épistémologique s'imposent aujourd'hui à la recherche scientifique orientée vers le travail sur des corpus numériques. Certes, chaque plateforme virtuelle présente une écologie spécifique (Paveau, 2013a, 2013b) qui oriente un rapprochement différent tant de l'objet que du corpus. Ainsi, l'environnement Facebook (FBK), une surface essentiellement multiforme suggère un regard qui puisse saisir son hétérogénéité sémiotique et énonciative. Dans ce travail, nous visons, d'abord, une redéfinition de la notion de corpus comme « matrice du sens » (Mayaffre, 2011 : 11) permettant de focaliser les enjeux scientifiques que la conception des corpus numériques issus du Web 2.0, notamment, de FBK, entraîne; ensuite, la description de certains concepts méthodologiques et épistémologiques fondamentaux – linéarité, technodiscours, sérialité, réticularité – aidant à l'élaboration ainsi qu'à la gestion de corpus FBK.

**Mots-clés :** corpus numériques, Facebook, épistémologie, réticularité

### Corpus web 2.0: algunos desafíos metodológicos y epistemológicos

#### Resumen

Numerosos desafíos de orden metodológico y epistemológico se imponen actualmente a la investigación científica orientada al trabajo sobre corpus digitales. En efecto, cada plataforma virtual presenta una ecología (Paveau, 2013ab) específica que promueve una aproximación diferente tanto del objeto como del corpus. Así, el entorno Facebook (FBK), un espacio esencialmente pluriforme sugiere una mirada que pueda aprehender su heterogeneidad semiótica y enunciativa. En este trabajo intentamos, primero, redefinir la noción de corpus como *matriz de sentido* (Mayaffre, 2011:11) de modo de focalizar los desafíos científicos que la constitución de corpus digitales nacidos de la Web 2.0, en particular, de FBK, acarrea; luego, describir ciertos conceptos metodológicos y epistemológicos fundamentales – linealidad, tecnodiscurso, serialidad, reticularidad – que ayudan tanto a la elaboración como a la gestión de corpus FBK.

**Palabras clave:** corpus digital, Facebook, epistemología, reticularidad

## Corpus Web 2.0: some epistemological and methodological challenges

### Abstract

Numerous challenges to methodological and epistemological oriented digital corpus work on scientific research is currently imposed. In effect, each virtual platform presents *ecology* (Paveau, 2013ab) specifies that promotes a different approach both the object and the corpus. Thus the Facebook environment (FBK), a space essentially suggests a multifaceted look that can grasp its semiotics and declarative heterogeneity. In this paper we try, first, to redefine the notion of corpus *sense matrix* (Mayaffre, 2011: 11) so as to focus the constitution scientists born digital corpus of Web 2.0, in particular, FBK, brings challenges; then describe certain methodological and epistemological fundamentals - linearity, tecnodiscurso, seriality, reticularity - that help both the development and the management of FBK corpus.

**Keywords:** digital corpus, Facebook, epistemology, reticularity

### Introduction

Les pratiques discursives écrites ont connu une évolution certaine qui a fait émerger des rapports sociaux, inscrits dans des modèles culturels et socio-économiques particuliers, identifiés à partir des progrès technologiques des TIC. Avec le développement d'internet et du Web social, la recherche scientifique orientée vers le travail sur des corpus numériques se voit assigner une tâche aussi incontournable que nécessaire : resituer les problèmes d'élaboration de corpus au sein de cette évolution en leur donnant un statut spécifique au cœur des nouveaux enjeux méthodologiques et épistémologiques. Certes, aborder les problématiques liées au recueil des données, à leur exploitation et à la constitution des corpus, quelle que soit leur nature, entraîne, à des niveaux différents de la démarche analytique, nombre de difficultés surtout sachant l'effort qu'implique, dès le début, la conception des corpus numériques issus du Web 2.0, notamment, de Facebook (FBK). De fait, bien que le caractère numérique soit commun à tous les supports nés du Web, chaque plateforme présente une écologie (Paveau, 2013a, 2013b) particulière et un mode de structuration des matériaux précis, ce qui oriente, par ailleurs, un rapprochement différent tant de l'objet que du corpus – même pour des corpus textuels non-numériques –. Ainsi, l'espace virtuel inhérent au réseau social FBK, une surface essentiellement multiforme, nécessairement polyphonique et intertextuel, suggère un regard qui puisse rendre compte de son hétérogénéité sémiotique et énonciative.

C'est pour cela que, dans ce travail, nous proposerons une double réflexion, non exhaustive, visant, d'abord, une redéfinition de la notion de corpus comme *matrice du sens* (Mayaffre, 2011 :11) qui permettra de focaliser les enjeux scientifiques que l'étape d'élaboration de corpus numériques entraîne, d'autant plus qu'elle conditionne le traitement du corpus ainsi que la configuration des résultats obtenus ; ensuite, la description de certains concepts méthodologiques et épistémologiques fondamentaux – linéarité, technodiscours, sérialité, réticularité – aidant à l'élaboration ainsi qu'à la gestion des corpus FBK.

### 1. Élaborer des corpus textuels : quelques contraintes générales

Adopter une perspective convenable ne semble pas être si évident. Il peut s'agir, par exemple, de fonder les décisions méthodologiques et épistémologiques sur la base des contraintes valables de façon de pouvoir expliquer les problèmes liés au genre, au support technique, au code, aux langues, entre autres. Quoi qu'il en soit, le défi est, avant tout, de justifier les découpages des ensembles collectés, puis, de savoir qu'ils doivent être en rapport avec la recherche d'objectifs précis appuyés en même temps sur des objectifs scientifiques particuliers.

De manière générale, la constitution et l'équilibre des corpus dépendent de leur adéquation à un projet de travail, ce qui exige de la part de l'analyste, l'élaboration non seulement des techniques de construction, mais également d'un plan établissant leur traitement. D'après Elizabeth Marshman (2003 : 2), les critères les plus importants pour l'utilité du corpus comprennent le type de corpus et l'adéquation pour le projet visé. Or, le corpus, en tant qu'*objet* de la recherche, demande, a priori, un traitement spécifique d'autant plus que sa qualité détermine les résultats obtenus. Ainsi, il ne s'agit pas de recueillir des « données », de les décrire ou de les classer suivant la démarche entamée par la linguistique des corpus. Nous cherchons, en revanche, à substituer la catégorie « d'élément émergent » – plus pertinente pour la description des écologies web 2.0 – à celle de « données ». Cette dernière, signalant les éléments signifiants extraits du corpus qui permettent à l'analyste de gérer l'interprétation, semble restreindre en quelque sorte la portée de ce qui sera pour nous la *trace numérique* (Merzeau, 2009) – empreinte déposée en ligne lors d'une quelconque action – c'est-à-dire, un élément signifiant non pas « donné », mais « émergeant » de par les actions ou pratiques d'inscription en ligne.

Cette première étape de positionnement épistémologique, se situant en amont de l'analyse, nous fournit des éléments indispensables au traitement des corpus ainsi qu'à la description de la *matérialité* des éléments à repérer au sein du support FBK. C'est la raison pour laquelle, il est question, à nos yeux, de concevoir le corpus

comme une totalité dont les unités discursives sont mises en contraste grâce à des variables visant à faire ressortir leur sens. Il est admis que le plus enrichissant pour faire sens est de contraster les séquences de discours observées en fonction des variables externes et/ou internes (Charaudeau, 2009 : 54-55). Mais, dans le cas des univers comme celui de FBK, entièrement hétérogènes, comment pouvons-nous déterminer ces variables et en fonction de quels critères les précisons-nous ?

FBK, notre champ de discours, ne possède pas les mêmes caractéristiques que, par exemple, le champ discursif politique, religieux ou de la presse. Les stratégies de déconstruction des corpus traditionnellement utilisées dans les domaines où l'on travaille avec des corpus non numériques, ne semblent guère nous apporter des outils appropriés. À première vue, ce que l'on perçoit sur FBK, c'est la nature plurisémiotique et multiforme d'une plateforme virtuelle ou « terrain » réticulaire, « hypermorcelé » et déployé dans une multiplicité d'hyperliens ou de « points de fugue » coexistant en synchronie. Souhaitant plutôt illustrer cet endroit aussi particulier qu'attirant, on dirait qu'il évoque un « happening », c'est-à-dire, la « mise en place, non traditionnelle, d'une scène quelconque de la vie quotidienne », où il y a non seulement des évocations, mais aussi, et surtout, du spontané et de l'imprévu. En tout cas, tenant compte de ce panorama peu encourageant, il est nécessaire de réfléchir à des critères méthodologiques valables qui ne négligent pas les contraintes imposées par un champ de discours extrêmement complexe, difficile à appréhender d'une seule fois et à saisir tant en synchronie qu'en diachronie.

Une manière de s'y prendre, visant la construction/déconstruction d'un corpus issu de FBK – composé de différents espaces juxtaposés ayant une fonction et une organisation particulières – peut être la considération de sous-groupes existant à l'intérieur des surfaces morcelées. Par exemple, dans notre cas, il peut s'agir de se situer au niveau du profil (l'espace de l'aménagement des données personnelles), de la boîte destinée à la recherche d'« amis » ou de l'espace nommé *accueil* (le fil d'actualités publiées au jour le jour par tous les contacts de l'utilisateur en question). Nous croyons que ce point de départ aide à expliquer non seulement l'instabilité de la surface – au niveau discursif – mais aussi ce caractère dématérialisé qui lui est associé et qui établit une première grande différence avec les textes imprimés. Nous sommes consciente que notre réflexion ne peut aborder, dans le cadre de cette étude, la totalité de l'expérience numérique mais elle peut, pourtant, fournir quelques pistes sur l'inscription d'un sujet capté et contraint par l'affichage à l'écran, lesquelles pourraient être confrontées à d'autres études portant sur le reste des interfaces de FBK aussi bien que sur d'autres plateformes, ce qui justifie, dans nombre de cas et selon les buts entrepris, le choix d'un corpus ouvert.

## 2. Qu'est-ce qu'on observe ? Contraintes technodiscursives générales

L'étape de l'observation est habituellement développée en se posant trois questions clés (Quivy & Campenhoudt, 1995) : qu'est-ce qu'on doit observer ? Sur qui ? Comment ? Derrière cette apparente simplicité, il existe de nombreuses difficultés lorsqu'il s'agit d'un corpus FBK. Examinons de près ces problèmes :

- A. La première contrainte qui s'impose à tout chercheur voulant se plonger sur FBK, c'est en fait qu'il doit inévitablement créer un compte. Puis, il lui faut même aménager son profil pour, ensuite, avoir le droit de « partir à la recherche d'amis », ce qui veut dire, rassembler des contacts qui, eux aussi, lui fourniront d'autres contacts, et ainsi de suite. Enfin, s'il le veut, il peut poster des commentaires et participer à « la vie active » de ce mur ;
- B. Le chercheur, par ailleurs, n'a pas accès à la totalité d'utilisateurs du réseau FBK, mais seulement à l'ensemble de contacts ou amis qu'il a choisis. Cela met l'analyste face à une deuxième contrainte, la première étant celle de devenir « facebookeur », et qui s'impose par la dynamique de la structuration de l'interface car elle gère certains mouvements qui nous échappent. Par conséquent, le recueil se réalise sur la base d'un premier découpage de l'univers d'étude que le chercheur n'a pas volontairement prévu ;
- C. FBK nous met face à une surface où le plurilinguisme est à l'ordre du jour ;
- D. Les annonces publicitaires de toute sorte et les nouvelles sont aussi fréquentes et éparpillées en surface ;
- E. La surface est plurisémiotique : signes iconiques (photos, vidéos), linguistiques, sonores et gestuels ;
- F. Pour pouvoir recueillir l'ensemble de commentaires postés à la suite d'une intervention quelconque et lorsqu'ils sont nombreux, l'interface les « cache » au moyen d'un bouton sur lequel il suffit de cliquer pour voir apparaître cet ensemble mentionné. En réalité, l'interface répond à une configuration préétablie puisqu'elle intervient aussi à la constitution et distribution des traces tout au long de la surface. Nous voyons ici un élément intéressant sur lequel il faudrait s'arrêter ;
- G. Nous remarquons aussi que, d'habitude, non seulement chaque photo apparaissant sur le mur nous pousse à aller voir l'album qui est joint derrière elle, mais aussi on trouve affichée sur place la suite des premiers commentaires qui ont été postés dans l'accueil. Or, ces fragmentations, qui sont en fait fonctionnelles car opérées sur le mur par l'interface, nous aident à homogénéiser l'ensemble que l'on retiendra comme corpus (la suite des commentaires n'a pas été retenue), et de cette manière cela permet d'avoir une vision plus globale d'une totalité extrêmement hétérogène, impossible

de saisir d'une seule fois. Cela est inhérent à tout travail qui porte sur le numérique.

- H. Les « distracteurs » ou hypertextes « imposés » par l'interface : icônes, fenêtres, publicités, liens, etc., enfin, tout ce que nous appelons « les points de fugue » jouent un rôle important à l'heure de la segmentation. En réalité, ils ouvrent de nouveaux parcours et déroutent un peu le suivi des commentaires figurant dans l'accueil sous forme de liste.
- I. Une autre contrainte importante concerne la manière de considérer l'ensemble des discours produits, notamment sur le fil d'actualités. S'agissant des textes, des documents, même oraux, on peut déterminer si l'on travaillera sur les documents entiers (la liste complète des interventions postées telle qu'elles apparaissent sur l'accueil) ou sur quelques parties qui se définissent clairement comme étant le début, le milieu ou la fin (quelques échanges choisis en fonction des critères méthodologiques que nous précisons après). Mais, en principe, la longue liste de commentaires ou conversations que constitue l'accueil nous révèle, en revanche, une prémisse importante : cette production écrite constitue un *technodiscours natif* (Paveau, 2014) « se construisant en perpétuité ». Encore une fois, nous voilà face à des problèmes de découpage, puisque les catégories traditionnellement utilisées en linguistique de corpus ne sembleraient pas pouvoir caractériser cet ensemble numérique. En effet, connaît-on le début ou la fin de l'accueil ? À vrai dire, quand on s'y plonge, on n'est que dans « le milieu ». En tout cas, le début et la fin n'ont de sens que lorsque l'on ouvre le compte et on consulte l'accueil un jour quelconque. Dans la liste qui s'ouvre à nos yeux on distingue un début, qui ne dure que quelques minutes, voire quelques secondes, et une fin qui est toujours déterminée par l'interface. Comme nous pouvons voir, le début et la fin de l'ensemble des productions est si difficile à envisager – même à imaginer des échanges « disparaissant dans l'air » –, que la vision d'une totalité ou d'un produit discursif maîtrisable devient vague. Dans ces contextes numériques, un rapport tendu, tant au corpus qu'à l'objet s'impose car nombre de choix méthodologiques échappent aux décisions du chercheur à cause de la dynamique même de la plateforme FBK ;
- J. De plus, la fragmentation de cette liste infinie de commentaires qu'est l'accueil présente d'autres inconvénients encore moins maîtrisables par le chercheur que les précédents. Supposons que l'analyste, lorsqu'il décide d'ouvrir le compte et qu'il retrouve par hasard une suite de commentaires qu'il avait déjà recueillie, s'aperçoit aussitôt que cette unité conversationnelle a changé, ce qui arrive parfois d'une minute à l'autre. En effet, on constate que de nouveaux commentaires s'y ajoutent et, en conséquence, le statut formel

de l'unité se voit aussi modifié. Que faire alors ? Cette dynamique, échappant à tout contrôle de la part du chercheur, mais constitutive malgré tout de la surface accueil, marque dans les productions écrites des contours dans le temps et dans l'espace qui ne sont jamais clairs. On dirait que les limites des échanges répondent à une logique liée aux contraintes techniques. Face à ce panorama, on s'aperçoit donc que même les catégories de l'analyse conversationnelle nous interrogent d'autant plus que la structuration des échanges pose un problème méthodologique sérieux.

### 3. Émergence du sens : entre *parcours* et *matrice de sens*

Pour mieux circonscrire le corpus FBK dans ce qu'il a de spécifique, nous voudrions partager ici deux pistes que Damon Mayaffre (2011) nous fournit et qui visent des décisions méthodologiques centrales : la notion de *parcours* et celle de *corpus comme matrice*. L'idée de *parcours* nous oriente vers un mode particulier de « lire » l'objet empirique que l'on se donne. Cela veut dire que le *parcours*, au sens large, couvre nécessairement trois étapes imbriquées, liées dès le début de la recherche. Selon Mayaffre (2011 : 10) ce qu'il faudrait mettre en avant, c'est que le plan de travail doit « [...] avancer sur trois pieds et ne jamais séparer le commencement, la fin et les moyens<sup>1</sup> ; la chose, le procédé et les conclusions ». Le corpus comme *matrice du sens* (Mayaffre, 2011 :11-12) vise justement l'étude des formes empiriques et stabilisées du discours, c'est-à-dire, des corpus textuels ou (techno) discursifs qui englobent les performances linguistiques produites par les sujets quand ils s'expriment – des données attestées – et qui se matérialisent dans des genres. Autrement dit, il faut ne pas envisager le corpus comme « une base de données » à interroger car « le sens ne se cache jamais derrière une base », mais plutôt, de co-construire dans un *parcours* de lecture le sens qui naît d'une méthode se voulant moins « probatoire » – visant à prouver et à établir – qu'heuristique – visant à interroger et à interpréter l'objet.

### 4. Contexte, linéarité, sérialité et réticularité

Il est clair que la notion de contexte trouve une place importante dans ce raisonnement. En effet, Damon Mayaffre (2011 : 12) n'hésite pas à reprendre le « principe d'architextualité » lorsqu'il explique la notion de *matrice de sens*. L'auteur essaie de montrer que le corpus « informe les textes qui le composent ». Ainsi, l'analyste travaillera-t-il sur un « corpus de textes » défini comme « un objet construit de manière *ad hoc* par lequel le chercheur problématise sa recherche » et non pas sur « une base de textes (ressource matérielle dans laquelle le chercheur pourra puiser sa recherche) » (Mayaffre, 2011 : 12).

Comme nous pouvons observer, l'articulation local/global nécessite une redéfinition de la notion de contexte, car cette catégorie de base établit les conditions de production « co(n) textuelles » du sens pour la maïeutique. Il est opératoire pour nous d'admettre que le sens naît aussi bien en contexte qu'en corpus. À vrai dire, le contexte, de par sa nature, surtout quand il est associé à la situation socio-historique et aux conditions de production, est insondable pour entamer des études strictement linguistiques qui cherchent à établir les catégories d'analyse sur la base d'un corpus responsable de « donner » une certaine « corporalité » linguistique au contexte. Précisant un peu plus cette pensée, ce sur quoi l'auteur insiste, c'est sur l'idée que le corpus est conçu comme constituant un réseau de relations hiérarchiques dont la forme maximale est le contexte, ou bien, « la forme maximale formalisable du contexte » ainsi que du cotexte et de l'intertexte.

Néanmoins, il nous paraît pertinent de considérer que le corpus n'est pas que texte sur FBK, et qu'il constitue « un signe » faisant émerger des relations non seulement hiérarchiques mais en réseau, ce qui n'empêche pas pour autant qu'il soit un ensemble cohérent et cohésif aussi bien que le texte. Pourtant, ce dernier, est cohérent et cohésif « par essence » — il existe dans la société sans que l'analyste y intervienne —, tandis que le corpus l'est « par existence », car volontairement construit par hypothèse (Mayaffre, 2011 : 15-16). C'est la raison qui nous pousse à nous demander si, au cœur d'un corpus multiforme FBK, il existe une cohésion entre les traces matérielles observées dans les sous-groupes découpés selon les besoins. Allant plus loin, si on rassemble toutes ces traces éparpillées en réseau, ne construit-on pas un ensemble de relations complexes ? La forme ou le sens de l'objet ciblé changera selon la place où nous nous situerons, au sein du corpus dynamique, pour la regarder de plus près. Or, dans quel but portons-nous l'attention sur le fait que la cohésion-cohérence constitue la condition première favorisant l'interprétation du corpus aussi bien que du texte ? En quoi cela pourrait-il nous aider à exploiter mieux le corpus FBK ? Évidemment, il s'agit d'un rapprochement par contraste de deux catégories d'où l'on dégage que l'analyste travaille ou construit en fait la cohésion-cohérence de son corpus de manière hypothétique. Et, faute de ce caractère essentiel, le corpus manquera inévitablement de pertinence scientifique. Si l'on souhaite se plonger dans la conception de corpus à nature numérique, on est obligé de mettre en relief certaines remarques essentielles concernant la structure des catégories adoptées. Si d'après Damon Mayaffre « le corpus est fondamentalement un objet sériel, le texte est d'abord un objet linéaire », il ne nous reste qu'à prévoir l'influence qu'une telle distinction peut avoir sur la construction et interprétation de corpus qui émergent sur FBK.

Au moment d'isoler les « unités » d'analyse – supposons que l'on choisisse de retenir les conversations du fil d'actualité accueil – il nous sera en effet très utile de faire le point sur la structuration linéaire des commentaires. De fait, pour pouvoir interpréter ces conversations, il est juste de « soupçonner » qu'elles sont organisées linéairement, et que les parties qui intègrent l'ensemble accueil, ne peuvent être renversées sans changer la structure globale de la conversation. De plus, chaque échange posté, en dépit des difficultés déjà mentionnées *supra* concernant la reconnaissance d'un début, d'un prolongement et d'une fin dans la conversation, présente, à sa manière, une organisation spécifique. En tout cas, ce qu'il sera nécessaire de définir, c'est la nature de cette structuration qui naît dans des surfaces numériques. En revanche, un corpus basé sur la presque totalité d'échanges de l'accueil, bien qu'ils se manifestent comme une série partiellement chronologique et tronquée, remet en cause l'idée de linéarité de par l'entrecroisement des *posts* renvoyant à des usagers divers, ce qui complique assez le repérage des sources énonciatrices et leur suivi. En réalité, chaque échange posté, même dans une suite chronologique, n'est pas directement associé à celui qui le précède ou à celui qui le suit – il suffit de constater le changement thématique ou d'interactants pour s'en convaincre –. Peut-être devons-nous penser à une linéarité particulièrement inhérente à l'accueil.

Outre ces soucis mentionnés, il est vrai que « le corpus série » semble être un mode adéquat d'expliquer que le changement de place des échanges n'altère pas cependant la structure globale de l'accueil. Mais, il n'en est pas moins vrai qu'il se peut que le texte (chaque échange) dans ces surfaces virtuelles, et à la différence d'autres types de textes non numériques, présente un aspect sériel. Nous voilà donc au cœur des problématiques cruciales qu'il faut résoudre pour une efficacité herméneutique. Nous sommes de l'avis que, malgré les dichotomies – « suite continue versus série discontinue », « linéarité versus sérialité » – que Damon Mayaffre tente de mettre en évidence, quoique non définitives même pour lui, il existe un mode particulier d'entreprendre l'analyse de notre corpus : la réticularité.

Certes, les nouvelles lectures du numérique mettent en cause les principes de la linguistique textuelle, en particulier, la vision du texte linéaire, c'est-à-dire, le texte dans sa structuration séquentielle. À vrai dire, ce sont les nouveaux corpus numériques qui imposent des mouvements interprétatifs envisageant autrement la production de signes dans les surfaces virtuelles. Grâce à un repérage des signes combinant, à la fois, la vision linéaire et sérielle, une nouvelle cohérence-cohésion naît. Concrètement, il s'agit d'une lecture articulée dans sa dimension syntagmatique et paradigmatique qui facilite, d'une part, l'entrecroisement de deux visions : « objet linéaire, d'abord, le texte est aussi traversé de sérialité et de réticularité (...)

Objet sériel, en premier, le corpus est – ne serait-ce que par ce qu’il est composé de textes linéaires – traversé par la linéarité et la séquentialité » (Mayaffre, 2011 : 18) ; d’autre part, la conception de ce qu’on appelle *corpus réflexif* (Mayaffre, 2002, 2006, 2011) pour embrasser d’une seule fois la vision « architecturale » du corpus associée à une navigation, voire exploration hypertextuelle généralisée.

### 5. Propositions méthodologiques pour un *corpus réflexif* FBK

Tenant compte des problématiques exposées précédemment ainsi que des concepts abordés, nous sommes en mesure de pouvoir avancer quelques propositions méthodologiques et épistémologiques qui, loin de focaliser la dimension linéaire et/ou hiérarchique des corpus, ou bien, de mettre en œuvre une démarche logocentrée du traitement des corpus FBK – qui opère par extraction des matériaux linguistiques isolés de leur contexte de production/interprétation. –, embrassent, en revanche, une approche écologique (Paveau, 2013b) en consonance avec l’idée de corpus réflexif.

Tout en étant isolables de l’environnement, quelques composants de l’univers choisi, à l’occasion, les extraits de conversations émergeant sur l’accueil – à nature parfois entièrement linguistique –, se laissent mieux saisir par une démarche basée sur la prise de captures d’écran. En effet, cet ensemble d’images, focalisant différentes zones de l’interface, peut donner un aperçu plus complet de l’environnement technique authentique qui sous-tend les échanges visés. Le but principal de cette démarche destinée au traitement de corpus est donc de pouvoir incorporer à l’analyse les aires, les objets qui ne seraient pas visibles si l’on se bornait à la méthode d’extraction logocentrée. De cette façon, les captures nous permettent de tenir compte :

- A. Des actions menées en ligne par les usagers et d’en expliciter plus ou moins précisément les parcours hypertextuels possibles tout au long d’une surface ;
- B. Des rapports étroits qui existent entre langage et objets signifiants dans un support technique ;
- C. Des rapports de juxtaposition qui relient les éléments et les contenus composant la page écran ;
- D. Des distributions des hyperliens ;
- E. De la nature paratextuelle et discursive des typographies et des couleurs ;
- F. De la visualisation de la page telle qu’elle est perçue par les utilisateurs ;
- G. De l’existence de certains espaces cachés à l’intérieur de FBK et qui ressortissent par le biais du clic ;
- H. Des changements au niveau formel et structural enregistrés par l’interface

lors des actions réalisées en ligne. En fait, on peut capturer peu à peu, et à chaque instant, la suite des modifications se produisant de manière décalée ;

- I. De la nature des signes qui sont mieux visualisés par capture.

La construction de corpus FBK met en évidence l'hétérogénéité structurale de l'univers de discours choisi tant au niveau de l'interface qu'au niveau des échanges. Or, une telle complexité, inhérente aux discours créés sur le Web 2.0, devrait être abordée depuis une perspective qui permette d'expliquer l'émergence des phénomènes discursifs aussi complexes qu'on le croit. De plus, constituer un corpus a des répercussions sur la façon d'appréhender la manifestation, explicite ou implicite, de l'objet d'étude puisque c'est lui qui possède toute l'information que l'on cherche, et c'est nous qui devons en enlever les indices pertinents qui nous orienteront pour formuler des explications satisfaisantes. Ce rapport à l'empirique, entraîne un rapport au théorique qui marque de nouvelles perspectives tendant à envisager l'étude du numérique en focalisant sa spécificité. Enfin, c'est l'articulation entre théorie et pratique qui, d'une part, autorise la production de nouveaux faits, ce qui serait en fait de nouveaux objets d'étude ; d'autre part, influence les résultats et les perspectives de la recherche. Ce nouveau statut heuristique/formel du corpus et des technodiscours qui en découlent ne révèle que le besoin scientifique d'appréhender les nouvelles « corporalités » des pratiques discursives numériques. Si le texte est associé naturellement à un support « physique », le papier, qui non seulement déterminait un mode de saisie linéaire et séquentiel, de gauche à droite, de haut en bas, mais aussi, une interprétation particulière appuyée sur une cohérence-cohésion aussi précises, le numérique, en revanche, « rend plus évidente 'l'artefactualité' des technodiscours ». D'une certaine façon, le numérique semble certainement être à la source de nouveaux comportements lecteurs/interprétatifs qui fondent les bases d'une perception des parcours individuels dans l'affichage, la multiplication de niveaux d'étiquetage, de circulation d'information (par fichier joint, Google, etc.) sans limites. En conséquence, la catégorie texte, telle qu'on la comprenait traditionnellement en linguistique textuelle ne peut être reconsidérée qu'à l'intérieur des transformations numériques.

### **En guise de conclusion**

Examinant de près l'influence de la transformation numérique se développant de plus en plus rapidement au niveau de l'exploitation des corpus, nous concluons que de la même manière que l'on construit, à la faveur du numérique, notre corpus, incarné en tant que « matériau » soumis au traitement et la manipulation, il est légitime d'envisager la démarche de l'analyste comme un mouvement de quête des traces d'inscription des usagers, et que, ce faisant, il « circule » en surface, il

établit son parcours, il s'arrête là où il trouve des indices pertinents, il fouille les hypertextes, enfin, il construit son réseau de relations signifiantes lui permettant de faire ressortir le sens dont il a besoin pour expliquer les phénomènes qu'il aura ciblés davantage dans son projet de recherche. Nous soutenons enfin que, dans le cas qui est le nôtre, négliger les problèmes liés à la constitution des corpus numériques limite la portée aussi bien de l'efficacité du traitement heuristique que d'un rapprochement novateur de l'objet d'étude qui est, pour ainsi dire, imposé par les nouvelles conditions de production. Nous insistons, donc, sur l'idée que travailler le numérique implique pour le chercheur, d'abord, un regard autre sur les phénomènes abordés ; ensuite, la considération du lien étroit entre objet d'étude, corpus textuel numérique et positionnement méthodologique et épistémologique.

Nous avons voulu, dans ce court travail, mettre en relief le souci d'articuler les facettes constituantes du langage en rapport à un entourage contraignant qui réarticule le discours sur la scène d'énonciation et le couple à la technique, si bien que l'hétérogénéité couplée à la plurisémiotité de la Toile sont mises en évidence. Pour nous, l'hétérogène est au premier plan : on n'échappe guère à la prise en compte des irrégularités perçues à différents niveaux du discours, au relèvement d'échanges feints, à la focalisation d'une production de signes pluriels, au changement de sens des mots. C'est la raison principale qui justifie le fait que pour aborder le phénomène FBK, on ne peut pas faire l'économie de tous les aspects qu'il convoque, qui le constituent et complexifient. Bref, accorder à l'hétérogène une telle importance, c'est le postuler comme prémisse, c'est le juger comme condition d'existence du phénomène numérique. On comprend donc qu'un tel regard sur le numérique doive, en premier lieu, partir de l'hétérogène comme manifestation primordiale du discours FBK. En deuxième lieu, préciser les catégories pertinentes afin de réussir l'analyse visée. Enfin, le nouveau rapport à l'empirique dont nous avons parlé, doit permettre non seulement d'articuler la pratique et la théorie, mais aussi, surtout sur le plan épistémologique, de conduire à interroger l'importance de l'interdisciplinarité dans l'élaboration des corpus complexes. À la fin de cette modeste présentation, nous n'avons pu que mentionner les conditions théoriques et matérielles sur lesquelles il faudra inexorablement travailler mais non sans présenter, bien évidemment, les atouts majeurs qui à l'heure actuelle concernent le souci de commencer à explorer un terrain sur lequel il reste encore beaucoup à faire.

### **Bibliographie**

Charaudeau, P. 2009. « Dis-moi quel est ton corpus, je te dirai quelle est ta problématique », Université de Paris 13, CAD, *Corpus*, 8, « Corpus de textes, textes en corpus », 37-66. En ligne : <http://corpus.revues.org/1674?&id=1674&file=1> [Consultée le 3 mars 2014].

- Marshman, E. 2003. « Construction et gestion des corpus : Résumé et essai d'uniformisation du processus pour la terminologie ». Étude réalisée pour le groupe *Éclectik*. L'Observatoire de linguistique Sens-Texte, Université de Montréal. En ligne : <http://olst.ling.umontreal.ca/pdf/terminotique/corpusnormes.pdf>. [Consulté le 31 janvier 2015].
- Mayaffre, D. 2002. « L'Herméneutique numérique », *L'Astrolabe*.  
En ligne : <http://2010icles/art0031.htm>. [Consulté le 3 février 2016].
- Mayaffre, D. 2006. « Philologie et/ou herméneutique numérique : nouveaux concepts pour de nouvelles pratiques ? », BCL (UMR 6039), Université de Nice. En ligne : <http://www.revue-texto.net/Parutions/Livres-E/Albi-2006/Mayaffre.pdf>. [Consulté le 31 janvier 2014].
- Mayaffre, D. 2011. « Vers une herméneutique matérielle Numérique, *Corpus textuels, Logométrie et Langage politique* », Volume 1, Université de Nice - Sophia-Antipolis, UMR 6039, *Bases, Corpus et Langage*, Mémoire de synthèse présentée en vue de l'Habilitation à diriger des recherches. Directrice Sylvie MELLET. En ligne : [http://hal.inria.fr/docs/00/65/53/80/PDF/HDR\\_Mayaffre\\_SynthA\\_se.pdf](http://hal.inria.fr/docs/00/65/53/80/PDF/HDR_Mayaffre_SynthA_se.pdf) [Consultée le 20 septembre 2015].
- Merzeau, L. 2009. « Du signe à la trace : l'information sur mesure. *Traçabilité et réseaux* », *Hermès* n° 53, 23-29. *Une revue de l'Institut des sciences de la communication du CNRS (ISCC)*. En ligne <http://documents.irevues.inist.fr/handle/2042/31471>. [Consulté le 31 janvier 2014].
- Paveau, M-A. 2013a. « Technodiscursivités natives sur Twitter. Une écologie du discours numérique », *Epistémé*, pp. 139-176. En ligne : <https://hal.archives-ouvertes.fr/hal-00859064>. [Consultée le 20 septembre 2015].
- Paveau, M-A. 2013b. « Du contexte à l'environnement : une approche écologique du discours », conférence à la Journée Doscila Paris Diderot, 5 avril 2013, *La pensée du discours [Carnet de recherche]*. En ligne sur <http://penseedudiscours.hypotheses.org/11322> [Consultée le 20 septembre 2015].
- Paveau, M-A. 2014. « Les énoncés natifs du web : analyse du discours des réseaux sociaux numériques (twitter, facebook, Pinterest) ». En ligne <http://www.labeurb.unicamp.br/anexos/MAP-Conf.pdf> [Consulté le 4 juillet 2015].
- Quivy, R., Campenhoudt, L. V. 1995. *Manuel de la Recherche en Sciences sociales*, Coll. *Psyco Sup*. Paris : Dumond.

## Note

1. L'auteur souligne, de plus, que la méthode n'est pas entendue comme une sorte de *technè* ou *empeiria* mais « un mode heuristique d'accès au sens » (Mayaffre, 2011 :10).