

# Introduction

## I. LE NUMÉRIQUE : CONVERSION, TRANSFORMATION, CIVILISATION

Au début d'un article consacré à la trace, Louise Merzeau insiste sur la portée générale du numérique : « L'essor du numérique ne se réduit ni à une nouvelle codification des contenus, ni à l'introduction d'un nouveau canal de circulation. C'est une transformation environnementale, qui affecte les structures et les relations. Une telle mutation ne déstabilise pas seulement des usages et des objets. Elle remet en question les modèles conceptuels qui servent à les formaliser. » (Merzeau 2009 : 23). Milad Doueïhi n'hésite pas quant à lui à parler de « processus civilisateur » dans *La grande conversion numérique* (Doueïhi 2008). L'expression courante, objet d'un débat social interminable dans lequel on n'entrera pas ici, est *révolution numérique*. Quel que soit leur nom, *révolution*, *transformation* ou *conversion*, les actions et les effets du numérique sont là, l'usage des technologies numériques, d'internet et des objets connectés s'étant progressivement intégré à nos existences, tout du moins dans les aires culturelles, sociales et géographiques où les outils informatiques et les technologies numériques ont pu se développer; il ne faut pas oublier en effet que le numérique, comme la démocratie ou la sexualité, est une notion profondément située et ne souffre aucune universalisation.

Devant cette évolution civilisationnelle, Dominique Boullier s'inquiète de l'immobilisme des sciences sociales. Dans un entretien pour *Le Monde* en 2015, il estime que dans notre société de la haute fréquence, on assiste à un décrochage des sciences sociales au profit des GAFAM<sup>1</sup> qui embauchent des cerveaux pour « faire tourner

---

1. *GAFAM* ou plus récemment *GAFAM*, est un sigle qui désigne les « géants du web » : Google, Apple, Facebook, Amazon et Microsoft.

les machines », selon de nouvelles méthodes de travail fondées sur le volume, la variété et la vélocité. Il déplore que les sociologues, après les linguistes, soient dépossédés du traitement des données : « Il arrive aux sociologues ce qui déjà est arrivé aux linguistes et aux géographes. Ces dernières décennies, les informaticiens ont pris la main sur le traitement automatique de la langue, puis sur la géomatique. Désormais, ce sont les statisticiens, les informaticiens et les mathématiciens qui traitent et interrogent les données sur la société récoltées sur Internet » (Boullier 2015 : en ligne). Cette mention des linguistes est intéressante à bien des égards, et correspond à une situation observable également à propos des productions langagières numériques natives, on y reviendra plus bas. Pour remédier à cette dépossession, Dominique Boullier « plaide pour des sciences sociales de troisième génération, qui s'emparent de ces nouveaux phénomènes [numériques] pour les requalifier : il faut inventer des concepts, des outils et limites de validité sur ces nouvelles données, ces traces auxquelles nous n'avions pas accès » (Boullier 2015 : en ligne). Ce que ce chercheur dit de la sociologie est parfaitement valable pour les sciences du langage, et plus particulièrement l'analyse du discours, science sociale par excellence. Le travail présenté dans cet ouvrage est une réponse à cette nécessité d'inventer de nouveaux concepts, outils et limites pour rendre compte du fonctionnement des discours natifs d'internet dans une perspective qualitative et écologique.

## II. LES SCIENCES DU LANGAGE ET LE DISCOURS NUMÉRIQUE NATIF

Le discours numérique natif est l'ensemble des productions verbales élaborées en ligne, quels que soient les appareils, les interfaces, les plateformes ou les outils d'écriture. Il pose aux sciences du langage telles qu'elles sont pratiquées jusqu'à présent un certain nombre de problèmes qu'on ne peut résoudre qu'en pratiquant la remise en cause des modèles conceptuels mentionnée plus haut par Louise Merzeau.

## 1. Une autre nature du signe : du discours au technodiscours

Dans l'article précité, elle fait d'ailleurs ce constat sur le signe :

Le signe, tel que les SIC et la sémiologie le définissent, procède d'un acte d'énonciation doté de sens et (en partie au moins) d'intentionnalité. L'empreinte numérique, elle, est automatiquement produite à l'occasion d'un calcul, d'un codage ou d'une connexion, le plus souvent sans que le sujet en soit conscient. Au lieu d'articuler une face sensible (signifiant) à une représentation psychique (signifié), la trace assigne une signature invisible à un comportement informationnel, qui n'est pas toujours perçu comme tel (Merzeau 2009 : 24).

Cette remarque concerne évidemment de près les sciences du langage, et il est remarquable qu'elle vienne d'une autre discipline. Les sciences du langage n'ont pas encore fait ce constat et accusent un retard important sur la question des univers numériques et de leurs productions natives, tant sur le plan épistémologique que théorique et méthodologique. La plupart des rares travaux qui existent jusqu'à présent sur les discours natifs d'internet ou du web peinent à prendre en compte leur dimension technique, intégrée à leur nature langagière du fait de la programmation informatique qui structure les univers numériques ; ils restent logocentrés, c'est-à-dire axés sur la seule matière langagière, considérée dans sa définition saussurienne et dualiste (« la langue envisagée en elle-même et pour elle-même », selon la célèbre formule du fondateur de la linguistique moderne). Ces travaux<sup>2</sup>, qu'ils relèvent de la communication médiée par ordinateur (par exemple Anis 1998, Marcoccia 2013, Panckhurst 2006, 2007) ou de l'analyse du discours (par exemple Amadori 2012, Amossy 2011, Cunha 2014, Jackiewicz 2016), isolent en effet la matière langagière et discursive extraite de son environnement technologique informatique, pour retrouver la forme des extraits de corpus de l'analyse du discours traditionnelle ou les énoncés normalisés et mis en forme du traitement automatique

---

2. On ne mentionne ici que des travaux francophones ; les références seront élargies hors de la francophonie dans le cours de l'ouvrage.

des corpus, sur lesquels sont souvent mobilisées des théories et méthodologies prénumériques. Comme le dit justement Isabelle Pierozak, ce sont des travaux qui utilisent internet « *for corpus* » et non « *as corpus* » (Pierozak 2014).

Or, la spécificité des discours natifs d'internet est justement, entre autres, leur intense relationalité, c'est-à-dire leur intégration dans un réseau de relations algorithmiques qui en assure le fonctionnement et la circulation, en même temps qu'elle leur donne des traits linguistiquement inédits comme la cliquabilité sur le plan morpholexical ou l'imprévisibilité sur le plan discursif. La mise à l'écart de la machine, considérée comme un composant extralinguistique, amène à travailler sur les formes nécessairement stéréotypées de la langue et non sur les formes singulières, composites, mixtes, pleines des bruits et des moteurs du monde, des discours empiriques natifs des univers numériques. Les approches qui intègrent pleinement la machine et ses techniques sont rares et l'on citera prioritairement les recherches du groupe IMPEC (Interactions multimodales par écrans), reposant sur les trois principes de la médiation comme espace d'intersection entre corps, machine et langage, de l'affordance communicative et du design de l'environnement électronique (Develotte *et al.*, 2011). Le travail sur les discours numériques natifs implique en effet, non seulement de se défaire de la conception logocentrée du langage mais également d'une représentation anthropocentrée de la machine : dans leur travail sur l'écriture des machines, Cléo Collomb et Samuel Goyet décrivent bien cette représentation qui réduit la machine (ou le programme, le logiciel, l'application, etc.) à un outil, car, dans l'imaginaire des chercheurs, une machine fonctionne, mais n'agit pas, applique des règles mais n'écrit pas ; bref, elle est « neutralisée » (Collomb 2017, Goyet, Collomb 2016). Et effectivement, la linguistique semble neutraliser la machine pour reverser l'ensemble de la responsabilité de création langagière sur l'humain ; d'où une perspective logocentrée qui retombe sur ses pieds saussuriens, respectant le dualisme qui distingue le linguistique de l'extralinguistique<sup>3</sup>.

---

3. On retrouve cette défense dualiste à propos des machines parlantes au sens propre, comme Siri, l'application d'Apple. François Perea montre cependant

Or il est possible, soutiennent Cléo Collomb et Samuel Goyet, de sortir de la vision anthropocentrée de la machine comme de sa sacralisation techniciste, pour envisager une technologie qui ne soit pas une anthropologie (Collomb, Goyet 2015). C'est le sens des termes néologiques composés avec l'élément *techno-*, employés dans le travail de l'analyse du discours numérique pour rendre compte d'une évolution théorique nécessaire : parler de technodiscours, de technomot, de technosigne, de technogène de discours et de technographisme, c'est inscrire dans l'analyse une option théorique qui modifie l'épistémè mainstream des sciences du langage. C'est affirmer que les discours numériques natifs ne sont pas d'ordre purement langagier, que les déterminations techniques coconstruisent les formes technolangagières, que les perspectives logo- et anthropocentrées doivent être écartées au profit d'une perspective écologique intégrative qui reconnaisse le rôle des agents non humains dans les productions langagières. Les sciences de l'information et de la communication parlent de technosémiotique depuis les années 1990, pour signifier cette prise en compte des dispositifs techniques dans l'analyse de la construction du sens ; le terme est naturalisé et repris par les chercheurs qui travaillent actuellement sur le web 2.0. Il n'est pas question d'hypertrophier le techno- et de tomber dans la mythologie de la machine toute-puissante destituant l'humain de son statut de sujet ; mais il semble hors de propos de conserver une définition homogène et exclusive du langage, comme produit de la seule faculté humaine contrôlé par l'intentionnalité du sujet. Parler de technodiscours, d'élément technolangagier ou d'analyse technolinguistique, c'est donc s'inscrire dans une pratique écologique et postdualiste de la linguistique.

---

que le dialogue entre l'humain et la machine Siri relève davantage de la collaboration que de la distinction dualiste, l'humain adoptant des formes technolangagières qui l'éloignent du langage dit naturel (Perea 2016).

## 2. Les acquis des sciences de l'information et de la communication

La perspective construite par l'analyse du discours numérique s'inscrit dans ce mouvement de critique de l'anthropocentrisme, impliquant, en sciences du langage, un logocentrisme. Et si l'on cite ici deux jeunes chercheurs en sciences de l'information et de la communication, ce n'est pas un hasard : cette discipline s'intéresse depuis longtemps (en France, une vingtaine d'années) aux discours numériques natifs sous d'autres étiquettes comme *écrits d'écran* ou *écriture hypertexte*, et elle a donc une avance considérable (voir les travaux d'Emmanuel Souchier et Yves Jeanneret notamment). Pour des raisons qui ne sont pas toutes scientifiques, et qu'il faudra bien expliciter un jour, les deux disciplines ont été et restent encore aujourd'hui séparées : une sorte de barrière s'est dressée entre elles, au détriment des sciences du langage en ce qui concerne le numérique. Les acquis des sciences de l'information et de la communication sur l'écriture numérique, l'énonciation éditoriale, les formats d'écriture et de publication, les différentes sémiotiques des productions natives, le rôle des machines, sont des préalables nécessaires à l'élaboration de théories et de méthodes linguistiques pour l'analyse des discours. Les propositions faites dans cet ouvrage s'y appuient constamment, dans un souci de cumulativité (pour éviter la réinvention de l'existant) et de validité scientifique (une analyse linguistique qui ne tiendrait pas compte de la nature technocommunicationnelle des énoncés passerait à côté de l'essentiel). Elles s'appuient également sur les acquis également anciens d'une autre discipline séparée de la linguistique, mais peut-être moins complètement, la littérature, qui fournit une réflexion sur l'hypertexte et les textualités numériques depuis les années 1990 également (chez Christian Vandendorpe, Jean-Pierre Balpe, Serge Bouchardon, Alexandra Saemmer par exemple).

## 3. Questions posées à la linguistique

Les discours numériques natifs posent des questions à la linguistique comme discipline, en interrogeant ses principes, ses méthodes et ses objets.

C'est d'abord la nature même du langage et de ses manifestations qui est interrogée : en ligne, ce ne sont pas vraiment les scripteurs et locuteurs qui écrivent et qui parlent mais, pour aller vite, les machines et leurs programmeurs qui permettent que des productions langagières fruits de l'intentionnalité de sujets soient performées et acquièrent une existence (Herrenschmidt 2007). Dominique Cotte parle à ce propos de « co-énonciation technologique » (Cotte 2004 : 111). Cet état de fait concerne absolument tous les discours numériques natifs, des plus linéaires au plus délinéarisés par des liens hypertextes, des technomots (hashtags, URL), des images, des sons, etc. Les formes produites sont en l'état actuel de la théorie linguistique, non analysables : que peut dire la description linguistique d'une URL, d'un hashtag, d'une pancarte numérique, d'une image macro, sauf à séparer les éléments selon leur sémiologie (texte d'un côté, image de l'autre, et code d'un troisième) et à défaire par conséquent ce qui fait justement leur spécificité, c'est-à-dire leur coconstitution structurelle ? Pour en rendre compte, on parlera ici d'éléments composites ; la production langagière *à* la machine est en fait une production *de* la machine et il s'agit là d'une évolution inédite dans l'histoire du langage dont les sciences du langage doivent se saisir.

Les discours natifs du web sont par ailleurs relationnels : l'architecture du réseau fait qu'ils sont tous matériellement interreliés, entre eux et à leur énonciateur, ce qui leur donne des propriétés particulières comme leur investigabilité (tout énoncé en ligne peut être cherché et trouvé *via* des outils de recherche comme les moteurs) et leur idionuméricité (tout énoncé du web en ligne possède une forme unique et subjective, déterminée par les paramètres de navigation, de sociabilité, de lecture et d'écriture de l'internaute). La linguistique possède dans ses bagages théoriques la notion de dialogisme et il serait tentant de l'appliquer à la relationalité des discours numériques natifs ; mais en ligne, la relationalité est matérielle et automatique, et ne dépend pas de marques d'intertextualité ou d'analogies issues de la compétence interprétative de l'analyste. Encore une fois, il faut faire avec la machine, et s'équiper pour rendre compte d'énoncés qui sont des liens, pris dans des circuits de collecte, de traitement et de redocumentarisation automatiques et, au sens propre, innombrables :

la relationalité a en effet pour conséquence une démultiplication hors de mesure de la circulation des énoncés en ligne.

Des questions précédentes découlent directement celle de la constitution des corpus. Comment constituer des corpus d'analyse du discours sur un terrain où les énoncés sont innombrables, interreliés, subjectifs et composites? Comment passer des corpus traditionnels de l'analyse du discours, à partir de documents imprimés ou d'enregistrements stabilisés et objectifs (au sens où ils présentent la même forme pour tout récepteur), à des corpus subjectifs, ouverts et évolutifs? Comment éviter la tentation du corpus logocentré et l'application de théories prénumériques tout en mobilisant le corpus de théories et de méthodes des sciences du langage? Les réponses à ces questions passent par une connaissance approfondie des univers discursifs numériques et l'élaboration de méthodes permettant de respecter l'écologie du web.

### III. TERRAINS ET MÉTHODES DE L'ANALYSE DU DISCOURS NUMÉRIQUE

L'analyse du discours numérique telle qu'elle est présentée ici constitue une analyse qualitative des discours natifs d'internet, en particulier du web 2.0, qui implique un certain nombre de choix méthodologiques.

#### 1. Internet et le web

Une précision s'impose sur la différence entre ces deux terrains, la confusion étant très fréquente chez les chercheurs en linguistique : rappelons qu'internet (*Inter Network*), apparu à la fin des années 1960, est un réseau qui relie des ordinateurs entre eux au niveau mondial, ce réseau proposant plusieurs services, le partage de fichiers, la messagerie instantanée, la téléphonie, l'envoi de courrier électronique et le web. Le web (*World Wide Web*) inventé en 1989-1990, est donc un service, une application d'internet, et ne coïncide pas avec lui. Il a une histoire marquée par des évolutions structurelles que l'on balise habituellement par des chiffres : le web 1.0 ou web statique, déployé dans les années 1990, connecte des informations et repose



sur le système « push » de distribution de l'information (c'est le web des portails d'information et des forums); le web 2.0 ou web social ou participatif, apparu au début des années 2000, connecte des personnes et repose sur l'interaction multi-agents (c'est le web des réseaux sociaux et du partage multimédiatique); le web 3.0, web des données ou web sémantique, qui émerge au début des années 2010, repose sur la curation, c'est-à-dire la collecte et l'organisation des données, organise le web en gisement de données grâce à des métadonnées et privilégie les connexions mobiles; on parle actuellement de l'émergence à l'horizon 2020 d'un web 4.0, web intelligent ou métaweb, qui intégrerait une dimension connectée à l'ensemble des éléments de notre environnement de vie. Ces différentes réalités, internet ou les différents webs, ne constituent pas des terrains équivalents pour la linguistique car la communication, l'interaction et la publication de contenus langagiers n'y ont pas les mêmes caractéristiques et ne permettent donc pas les mêmes formes langagières et discursives. C'est la raison pour laquelle le logocentrisme de certaines approches ne permet pas de rendre compte de tous les aspects de la communication en ligne : rassembler et classer les données langagières selon les seuls critères linguistiques, en n'intégrant pas la dimension socio-technique des interfaces et du codage informatique ni les modalités d'éditorialisation des contenus implique une homogénéisation et une normalisation des contenus qui passe à côté de traits spécifiques.

On se concentrera dans cet ouvrage sur les discours numériques natifs du web 2.0 majoritairement francophone (certains exemples concernent l'anglais), avec quelques incursions dans le web 1.0 ou le web des données. Les terrains d'analyse seront donc essentiellement l'immense continent des réseaux sociaux numériques (notamment les plus connus et pratiqués, Twitter, Facebook, Youtube), la blogosphère, la presse en ligne, mais aussi certains sites et forums 1.0, certains outils de fabrication de discours en ligne. Les énoncés envisagés dans cet ouvrage sont tous natifs d'internet, c'est-à-dire produits directement en ligne et non portés en ligne à partir d'une numérisation hors ligne ou composés sur un téléphone hors connexion. La réflexion étant d'ordre théorique et non applicative, le travail se fait à partir d'exemples, et non de corpus constitués.

## 2. Position du chercheur et présentation des exemples

La connaissance des objets de la recherche est une vieille question en sciences humaines et sociales, qui croise celle de l'observation participante, de la réflexivité du chercheur, du paradoxe de l'observateur, de son engagement ou de son immersion, du statut de ses données biographiques, etc. Tous les objets de recherche ne réclament pas forcément leur connaissance interne mais l'étude des discours numériques natifs s'en passe difficilement, sauf à rester dans une approche logocentrée et anthropocentrée. Il semble en effet nécessaire que le chercheur soit aussi un usager d'internet et du web, ait une connaissance minimale des interfaces techniques et du fonctionnement de la machine. De nombreux phénomènes technodiscursifs ne s'identifient et ne s'analysent qu'à partir de cette connaissance pratique : la nature d'un lien hypertexte, le fonctionnement d'un mot cliquable, l'idionuméricité et l'imprévisibilité des contenus, les formats et les genres de texte. À défaut, la description emprunte des analogies avec des phénomènes non ou prénumériques dont certaines seront mentionnées dans l'ouvrage.

La perspective écologique impose de présenter les exemples dans leur environnement natif et l'idéal serait évidemment de pouvoir travailler avec un navigateur ouvert, ce que ne permet évidemment pas la publication hors ligne. On a donc choisi de présenter les exemples sous forme de captures d'écran, ce qui est le minimum écologique nécessaire, même si ce procédé semble figer les données technodiscursives ouvertes et mobiles, et les objectiver : les captures d'écran sont elles aussi issues de la subjectivité de l'internaute-analyste et il faut les considérer comme des données subjectives.

## IV. ORGANISATION DE L'OUVRAGE ET CONVENTIONS GRAPHIQUES

Cet ouvrage a été commencé sous la forme d'un dictionnaire sur un carnet de recherche, *Technologies discursives* (<<https://technodiscours.hypotheses.org>>), forme qui a été conservée dans le présent ouvrage. Les notions sont donc présentées sous forme d'entrées alphabétiques construites à partir d'une définition, la table des

matières et un index des notions permettant de se repérer dans leur contenu. La liste d'entrées présentée n'est évidemment pas exhaustive et l'ouvrage ne prétend pas faire le tour des questions posées par les discours numériques natifs, mais en proposer un cadre d'analyse orienté par l'approche écologique.

Dans tous les exemples cités, hors illustrations, les graphies d'origine sont conservées. Les nombreux anglicismes de la recherche sur internet et le numérique en général ont été maintenus, la traduction française étant parfois maladroite (le malheureux *mot-dièse* n'est pas employé par exemple) et souvent inutile, les mots anglais s'étant vite lexicalisés. Les citations en anglais ne sont pas traduites, le statut de lingua franca de l'anglais scientifique ne le demandant plus. L'écriture épïcène a été écartée à regret (sauf dans les remerciements) pour des raisons d'économie graphique et de lisibilité, le nombre de néologismes, d'emprunts et de graphies particulières étant déjà très élevé dans l'ouvrage.