Corpus

8 | 2009
Corpus de textes, textes en corpus

Tabularité : des textes aux corpus

Marie-Laure Florea

Édition électronique
URL : http://journals.openedition.org/corpus/1792
ISSN : 1765-3126

Éditeur
Bases ; corpus et langage - UMR 6039

Édition imprimée
Date de publication : 15 novembre 2009
Pagination : 177-196
ISSN : 1638-9808

Référence électronique
Marie-Laure Florea, « Tabularité : des textes aux corpus », *Corpus* [En ligne], 8 | 2009, mis en ligne le 27 mai 2011, consulté le 07 septembre 2020. URL : http://journals.openedition.org/corpus/1792

© Tous droits réservés

Tabularité : des textes aux corpus

Marie-Laure FLOREA
ICAR (UMR 5191), Université Lumière Lyon II – CNRS

Introduction

Histoires dont vous êtes le héros, « texte mosaïque »[1] des magazines de presse, calligrammes – dans lesquels la disposition du matériel signifiant sur la page fait sens –, publicités dont le message est véhiculé par plusieurs systèmes sémiotiques distincts, hyperfiction[2] : nombreux sont les textes qui ne se lisent pas de façon uniquement linéaire, mais dont le parcours de lecture est multiple, pluriel, ne se déroulant pas selon un fil unique. En effet le texte écrit d'aujourd'hui n'a plus grand-chose à voir avec celui d'hier, qui se contentait de reproduire sur un plan spatial la linéarité temporelle de l'oral (où les sons, mots et phrases se succèdent inévitablement). Et pour cause : le canal visuel offre des potentialités dont le canal auditif ne dispose pas.

Et pourtant, force est de constater que texte et corpus continuent d'être envisagés sous leur aspect uniquement linéaire. Or évincer cette dimension non-linéaire du texte revient à n'avoir de ce dernier qu'une vision partielle et partiale, relayée par la présentation traditionnellement linéaire des corpus. Ceci n'est pas sans poser de multiples problèmes théoriques et méthodologiques au chercheur confronté à des textes non exclusivement linéaires : comment intégrer dans un corpus des éléments sémiotiquement hétérogènes (images, infographies...) ? Comment gérer au sein d'un corpus

1 Selon l'expression de Mc Luhan [1967 : 42]

2 Œuvre de littérature électronique interactive où le lecteur choisit lui-même son parcours, par exemple *Afternoon, a Story*, première hyperfiction publiée en 1987 sur disquette par Michael Joyce, *20% d'amour en plus* de François Coulon (publié sur CDRom), *Vaisseaux fantômes* de Renaud Camus (paru sur Internet).

Corpus n°8 « Corpus de textes, textes en corpus »
(2009), 177-196

l'hétérogénéité des statuts entre les différentes composantes du texte (titres, chapeaux, intertitres, corps de l'article, nom d'auteur...) ? Comment tenir compte des relations entre les divers éléments qui composent un texte complexe ? À quelle échelle parler de texte ?

Ce sont ces questions très pragmatiques qui m'ont initialement amenée à m'interroger sur la gestion des discontinuités dans les textes et les corpus. Mes recherches portent en effet sur les nécrologies[3] dans la presse et mon corpus[4] comporte plusieurs hyperstructures[5]. En préparant le tableau qui fait la synthèse des articles de mon corpus, je me suis demandé comment présenter ces hyperstructures : la nécrologie hypertextuelle dans son ensemble devait-elle faire l'objet d'une entrée unique ou chaque article qui compose l'hyperstructure devait-il bénéficier d'une entrée distincte ? C'est ainsi la phase de mise en corpus qui incite souvent à se pencher sur « le statut et la perception du texte antérieurement défini comme unité globale, comme tout cohérent et auto-suffisant », pour reprendre les termes de l'appel à communications de ce numéro.

Ce sont ces difficultés auxquelles je me suis heurtée lors de la mise en corpus qui m'ont amenée à vouloir proposer ici une théorisation du fonctionnement du type de texte particulier auquel appartient l'hyperstructure journalistique, et que j'appellerai texte tabulaire. Je désigne par texte tabulaire un texte composé de plusieurs modules, ayant chacun une autonomie relative mais étant interdépendants les uns des autres, regroupés sur un espace matériel borné. Il existe en effet

3 Par nécrologies, j'entends les articles retraçant la biographie d'une personnalité récemment décédée, et non les faire-part de décès publiés par les familles.

4 Corpus composé de quelque cinq cents nécrologies, recueillies (de façon non sélective) dans l'ensemble des titres de la presse quotidienne nationale française entre avril et juin 2006.

5 Que Lugrin définit comme un « élément de structuration de l'information, intermédiaire et facultatif, situé entre le journal et l'article […], formé d'un ensemble d'articles et d'images graphiquement regroupés et complémentaires, bornés à la limite matérielle de [...] la double page » [Lugrin 2001 : 69]

diverses études qui analysent cette tension entre unité et discontinuité dans certains textes, telle cette analyse menée par René Audet [2000] sur le recueil de nouvelles (dont il tente de montrer qu'il forme une totalité signifiante et non une suite de textes discontinue), ou les travaux de Lugrin [2000] et d'Adam et Lugrin [2000, 2006] sur les hyperstructures journalistiques. Cependant, à ma connaissance, il n'existe pas d'étude synthétique sur le fonctionnement du texte tabulaire, en dehors d'un genre ou d'un type de discours déterminé. C'est par conséquent ce que je me propose d'esquisser ici.

L'objectif de cet article est double : il s'agit d'une part de décrire, dans une perspective croisée d'analyse du discours et de linguistique textuelle, le fonctionnement de ce type de texte particulier, mais loin d'être marginal, et, par ailleurs, d'appliquer ces perspectives au fonctionnement des corpus, qui restent aujourd'hui calqués sur le modèle du texte linéaire.

Je commencerai par revenir sur l'opposition entre linéarité et tabularité, afin de distinguer deux régimes de textualité distincts. Je me pencherai ensuite plus particulièrement sur le régime de textualité tabulaire, afin d'en décrire le fonctionnement original. Je terminerai en tentant de renouveler l'approche des corpus, à la lumière des éléments de description du texte tabulaire que j'aurai dégagés, afin de permettre aux corpus d'en refléter la textualité particulière.

1. Linéarité & tabularité

1.1 État des lieux : le primat du linéaire

Aujourd'hui, la conception du texte demeure marquée par le primat du linéaire : un petit aperçu des définitions du texte montre à quel point la linéarité imprègne la représentation du texte, aussi bien chez l'analyste que chez le lecteur. En effet, la définition du texte proposée par la grammaire de texte :

> Séquence bien formée de phrases liées qui progressent vers une fin. [Slatka 1985 : 138]

diffère assez peu des définitions plus modernes que propose l'analyse du discours :

> Suite linguistique autonome, orale ou écrite, produite par un ou plusieurs énonciateurs dans une situation de communication déterminée. [Maingueneau 1996 : 81]

> Suite d'énoncés oraux ou écrits posés par leur producteur – et destinés à être reconnus par leur(s) destinataire(s) – comme un ensemble cohérent progressant vers une fin et parvenant à constituer une complétude de sens. [Barbéris, *in* Détrie, Siblot, Verine 2001]

Cette notion de séquentialité se retrouve également dans les définitions du texte que proposent des dictionnaires non spécialisés :

> La suite des mots, des phrases, qui constitue un écrit ou une œuvre (écrite ou orale). [NPR]

> Suite de signes linguistiques constituant un écrit ou une œuvre. [TLFI]

Ce primat du linéaire s'explique probablement en partie par des raisons historiques : les premiers textes alphabétiques furent écrits en boustrophédon, constituant un fil d'écriture ininterrompu allant de gauche à droite puis de droite à gauche sur la ligne suivante : les origines du texte ont sans aucun doute marqué la représentation que l'on en a aujourd'hui. Par ailleurs des raisons idéologiques s'ajoutent probablement à ces raisons historiques : le livre est considéré aujourd'hui comme le texte par excellence, or les livres invitent la plupart du temps à une lecture linéaire.

1.2 Deux régimes de textualité

Cependant bien des aspects du texte viennent remettre en question la centralité de la linéarité. Un texte est en effet traversé par de nombreuses discontinuités qui viennent segmenter le fil linéaire de la lecture : ces discontinuités peuvent être marginales (en cela qu'elles affectent un texte qui se lit majoritairement dans un ordre linéaire déterminé) : il s'agit du titre, du nom d'auteur, des notes, du sommaire... Elles peuvent aussi être constitutives de la textualité, dès lors qu'elles affectent la structure du corps du texte lui-même, qui se trouve subdivisé en plusieurs modules textuels.

Cette distinction entre les discontinuités marginales et structurelles me permet de distinguer deux régimes de textualité différents, qui sont autant de façons d'agencer le matériau langagier pour former un texte : le régime de textualité linéaire et le régime de textualité tabulaire.

Le régime de textualité linéaire caractérise les textes composés d'une suite de mots progressant globalement de façon linéaire du début vers la fin. La relation d'ordre entre les différents éléments du texte est très forte et demeure même lorsque la lecture transgresse la linéarité prévue par l'auteur : ainsi, même si on commence à lire un roman par la fin, on a conscience de ne pas respecter l'ordre conventionnel de la lecture et on sait que ce qu'on lit est la fin de l'ouvrage.

Dans le régime de textualité tabulaire, l'appropriation du texte par le lecteur se fait de manière fondamentalement différente, n'étant plus balisée par les notions de début, de milieu, de fin, mais par l'organisation spatiale du matériau signifiant. On peut considérer comme tabulaires les textes composés de modules plus petits, qui peuvent être sémiotiquement hétérogènes mais à dominante linguistique, et dont la combinaison forme un texte, c'est-à-dire un tout qui représente plus que la simple somme des parties qui le composent. C'est le cas des hyperstructures journalistiques, des recueils littéraires (de poèmes, de contes, de nouvelles), des recettes de cuisine, des tracts, des posters scientifiques... Cette configuration textuelle originale permet d'entrer dans le texte par plusieurs portes et pas seulement de le parcourir depuis le début jusqu'à la fin.

Dans les deux cas, les parcours de lecture demeurent balisés, mais selon des modalités différentes : dans un texte à dominante linéaire, ce sont les « points cardinaux » de la page qui guident la lecture, qui se fait de gauche à droite et de haut en bas. Dans un texte tabulaire, ce sont la topographie et la mise en page qui balisent la lecture. On sait en effet que le lecteur sera attiré par le centre plutôt que par la périphérie, par les pavés textuels courts plutôt que longs, par la taille des caractères, par la couleur... Il s'agit donc de deux modes d'accès au texte fort différents.

Il faut toutefois souligner que les notions de tabularité et de linéarité ne sont pas l'apanage exclusif des textes respectivement tabulaires et linéaires : en effet, ainsi que nous l'avons déjà vu, la plupart des textes linéaires sont affectés par des discontinuités dans la lecture. Parallèlement, tout texte tabulaire demeure marqué par une certaine linéarité, chacun de ses modules pris isolément étant organisé de façon linéaire.

1.3 Questions de terminologie

Pour mieux décrire ces deux types de textes, il semble essentiel d'enrichir la terminologie existante, particulièrement pauvre pour désigner les différents éléments qui composent le texte, et notamment le texte tabulaire.

Les textes linéaires peuvent être subdivisés, selon une distinction couramment admise, entre le corps du texte (par exemple la suite de mots qui compose un roman, du début à la fin de la narration) et le péritexte, qui depuis Genette [1987], désigne les marges du textes (titre, nom d'auteur, notes...) contenues dans les limites du volume. Le terme, bien que consacré par l'usage, est peu satisfaisant : par son étymologie, il relègue en effet hors du texte tout ce qui ne fait pas partie du corps du texte, alors qu' « il est nécessaire de considérer les éléments péritextuels comme des parties de l'unité complexe texte », ainsi que le propose Adam [1999 : 80]. Le préfixe laisse par ailleurs entendre que le péritexte serait secondaire par rapport au texte, or le péritexte est tout sauf périphérique et joue au contraire un rôle souvent central dans le fonctionnement du texte. Je n'ai malheureusement pas le loisir ici de me pencher plus avant sur cette question, souhaitant me consacrer au régime de textualité tabulaire, mais on voit à quel point la linéarité se laisse déjà interroger dans les textes linéaire, qu'il s'agisse des relations particulières que tisse le corps du texte avec le péritexte, ou de la disposition spatiale des éléments verbaux.

Dans un texte tabulaire, je propose d'appeler module chaque élément du texte qui, tout en ayant une certaine autonomie, entretient des relations étroites avec les autres modules qui composent le texte. Le texte tabulaire se lirait par conséquent à deux échelles : échelle du texte complet, constitué du corps du texte (l'ensemble des modules qui le composent) et

éventuellement d'un péritexte. Ce fonctionnement bipolaire serait reproduit à l'échelle du module textuel, lui-même composé d'un corps de texte et éventuellement d'un péritexte. On entrevoit déjà les relations multiples qui peuvent se tisser au sein de cette architecture complexe : des liens peuvent s'établir entre le péritexte et le corps du texte, entre le péritexte principal et le péritexte de chaque module et entre le corps de texte de chaque module.

Je choisis délibérément de ne pas parler d'hypertexte pour désigner les textes tabulaires, malgré une terminologie en vogue, pour deux raisons principales. Tout d'abord dans la littérature actuelle, l'hypertexte ne s'applique qu'à des textes électroniques. Je soutiens au contraire que l'essor du texte tabulaire, même s'il a sans aucun doute été favorisé par le développement du numérique, ne dépend pas du support : on trouve des textes tabulaires aussi bien sur des écrans que sur du papier.

Par ailleurs, dans la terminologie actuelle, hypertexte s'oppose habituellement à texte : il existerait donc deux formes différentes d'écriture, l'écriture textuelle et l'écriture hypertextuelle, comme le suggèrent Laufer et Scavetta [1992 : 5] :

> Un texte est une structure linéaire plus ou moins fortement hiérarchisée. Les éléments textuels, plus ou moins autonomes, sont reliés par des relations d'ordre. Un hypertexte est une structure de réseau : les éléments hypertextuels sont des nœuds reliés par des relations non-linéaires et faiblement hiérarchisés.

Je soutiens au contraire qu'il existe deux régimes de textualité distincts, et qu'un texte tabulaire est bel et bien un texte, au même titre qu'un texte linéaire. La différence se situe uniquement dans la configuration du matériau sémio-linguistique, mais pas dans la nature de l'objet. La notion d'hypertexte entrerait par conséquent de plein droit dans la réflexion sur l'intertextualité (dans la mesure où il s'agit d'une forme particulière de relation entre plusieurs textes distincts), mais non sur la textualité.

2. Les textes tabulaires

2.1 Question d'échelle

Avant de se pencher sur le texte tabulaire, il convient de justifier qu'il s'agit bien un texte. Cela ne va en effet pas de soi, et semble aller contre l'opinion couramment admise. L'échelle à laquelle joue le texte pose en effet problème dans les textes tabulaires : que doit-on considérer comme un texte, l'ensemble (à savoir la totalité d'un dossier de presse, ou d'un recueil de poèmes, par exemple), ou chaque module textuel pris isolément (donc chaque article du dossier, ou chaque poème du recueil) ?

La terminologie couramment utilisée pour désigner ce que je choisis résolument d'appeler texte tabulaire laisse supposer que l'échelle du texte joue au niveau inférieur, celui de ce que j'ai appelé module textuel. On parle en effet souvent de « multitexte », ce qui implique que le texte tabulaire soit de nature supratextuelle. On parle également dans le cas de la presse d'hyperstructure, ou d'ensemble rédactionnel, ce qui laisse à penser qu'il ne s'agit que d'une sorte de coquille vide (« structure ») qui ne présente une unité que sur le plan de la composition (« rédactionnel »).

Un élément supplémentaire qui plaiderait pour la nature supratextuelle de l'hyperstructure est l'échelle à laquelle le genre de discours est perceptible : dans son article de *Semen* 13 (dans lequel il étudie les hyperstructures journalistiques), Gilles Lugrin montre que l'essor des hyperstructures dans la presse s'accompagne d'un renforcement des genres de discours journalistiques (notamment par le biais d'une autodésignation des articles qui composent l'hyperstructure, « interview », « dépêche », « commentaire »...). Les limites du genre de discours se superposant habituellement à celles du texte, cela viendrait confirmer que le texte existe à l'échelle des articles qui composent l'hyperstructure.

Cela dit, considérer ces structures complexes comme « multitexte », même si cela comporte une certaine pertinence (et notamment dans le cadre de structures relativement longues, où le lecteur percevra souvent intuitivement l'échelle du texte au niveau de ce que j'ai appelé « module textuel »), n'épuise pas la question : il est indéniable que la structure dans son

ensemble possède également une unité, qui permet de la considérer comme un texte.

Il me semble d'ailleurs que la détermination de l'échelle à laquelle joue le genre ne va pas dans le sens de ce qu'a décrit Lugrin, dans de nombreux cas de textes tabulaires. Un poster scientifique, un tract électoral, une nécrologie hyperstructurelle : dans ces cas et bien d'autres, le genre de discours se lit au niveau du texte tabulaire dans son ensemble, et non au niveau de chacun des modules qui le composent. Or si l'on maintient que les frontières du texte se superposent à celles du genre de discours, cela valide le fait que les textes tabulaires sont bien des textes, et non des structures supratextuelles[6].

2.2 Autonomie et dépendance

Toutefois, si le texte tabulaire est bien un texte unique et non une somme de textes indépendants, la textualité s'y organise de façon fort différente de ce que l'on constate dans un texte linéaire. En effet, la tension traditionnellement conférée au texte entre continuité et progressivité, qu'on peut lire sous la plume d'un Adam par exemple :

> La textualité peut être définie comme un équilibre délicat entre une continuité-répétition, d'une part, et une progression de l'information, d'autre part. [Adam 1991 : 102]

6 Ceci n'invalide toutefois nullement les constatations de Lugrin, simplement je pense qu'il existe une confusion entre les genres de discours (considérés comme des formats de communication ritualisés) et les genres spécialisés, propres à un type de discours (comme ici le discours médiatique), qui ne sont pas des genres de discours, mais uniquement des moules codifiés dans lequel le texte vient se couler. On constate d'ailleurs qu'il existe de nombreux manuels de journalisme qui visent à enseigner au futur journaliste des genres codifiés (comme par exemple Agnès Y. (2002). *Manuel de journalisme. Écrire pour le journal*. Paris : La Découverte, ou Martin-Lagardette J.-L. (2003). *Le guide de l'écriture journalistique*. Paris : La Découverte) alors qu'il n'existe aucun livre qui permette d'apprendre à pratiquer les genres de discours, uniquement des travaux qui visent à les décrire. La distinction entre ces deux types d'ouvrages, les uns prescriptifs, les autres descriptifs, suffit à mon sens à montrer qu'on ne peut confondre genres journalistiques et genres de discours. Par conséquent, la position que je défends ici n'est nullement incompatible avec les conclusions de Lugrin.

est propre au texte linéaire, et elle est remplacée dans le texte tabulaire par une tension entre autonomie (de chaque module textuel) et dépendance (de chaque module par rapport à l'ensemble).

Cette tension repose sur l'éclatement du texte en modules textuels distincts, qui peut se faire selon des modalités différentes. L'éclatement peut en premier lieu se faire selon les différences de systèmes sémiotiques : une double page de manuel scolaire se subdivisera ainsi en plusieurs modules : un module linguistique généralement placé sur la page de gauche, qui constituera le « cours », et plusieurs modules d'illustration (photographies, infographies...) disposés traditionnellement à droite[7]. L'éclatement repose alors, en partie du moins, sur des critères sémiotiques, le linguistique à gauche, l'iconique à droite.

L'éclatement peut aussi se faire en fonction du contenu. C'est ce qu'on constate habituellement dans les nécrologies tabulaires, dans lesquelles les modules textuels correspondent souvent assez précisément aux séquences d'une nécrologie linéaire. Les nécrologies linéaires comportent généralement trois séquences : une brève séquence informative (destinée à annoncer la mort), une séquence narrative (qui vise à raconter la vie) et une séquence argumentative évaluative (qui rend hommage au disparu). Il existe une corrélation très forte entre cette organisation tripartite et l'agencement des modules dans les hyperstructures nécrologiques, qui comportent traditionnellement un article destiné à annoncer la mort (qui correspond à la séquence informative), un récit de vie et un encadré sur l'œuvre (qui reprennent la séquence narrative), une illustration (généralement une photographie du disparu à l'appui de la séquence narrative) et enfin un ou plusieurs articles d'hommages (qui correspondent à la séquence argumentative).

7 Comme le précisent les instructions suivantes : « De très nombreux manuels de l'école primaire, du collège comme du lycée, fonctionnent sur le modèle de la double page. Une des deux pages est consacrée à l'exposé des connaissances, en vis-à-vis l'autre page prolonge et illustre le cours, propose documents, expériences, exercices. La double page doit former un tout. ». Rapport de Dominique Borne sur le manuel scolaire : http://lesrapports.ladocumentationfrancaise.fr/BRP/994000490/0000.pdf

Ce parallèle montre que les nécrologies tabulaires sont une sorte d'éclatement des nécrologies simples, en fonction du contenu.

Enfin l'éclatement peut aussi reposer sur les différences de point de vue sur l'objet de discours : dans une recette de cuisine, chacun des trois modules traditionnels adopte un point de vue différent : la liste d'ingrédients adopte le point de vue de celui qui va faire des courses, la description de la préparation, de celui qui va confectionner la recette, et la photographie du résultat, le point de vue de celui qui va déguster le plat.

Cette liste (non exhaustive) des modalités de l'éclatement du texte en modules explique la tension entre autonomie et dépendance qui sous-tend les textes tabulaire : chaque module correspond à un regard original sur le référent et peut donc se lire de façon indépendante, mais c'est la complémentarité de ces regards qui donne sa teneur au texte.

2.3 Procédés de textualisation

Au-delà de cette complémentarité, l'interdépendance des modules textuels entre eux se construit par le biais d'un réseau de relations très dense qui se tisse entre eux. J'étudie ici ces phénomènes dans des nécrologies hyperstructurelles, mais on les retrouve dans d'autres types de textes tabulaires.

Le premier type de lien qui unit les modules d'un texte tabulaire concerne les renvois internes : il est en effet fréquent que le corps du texte d'un module invite le lecteur à regarder un module voisin, comme dans cette nécrologie d'André Mandouze dans *La Croix*, dont l'article principal comporte le passage suivant, qui côtoie un hommage de Gérard Depardieu au disparu :

> Une de ses dernières apparitions publiques aura consisté à prendre la parole dans les cathédrales, aux côtés de Gérard Depardieu (lire ci-contre).

La relation entre les modules peut également être liée à la mise en page : le texte tabulaire est en effet circonscrit dans un ensemble clairement délimité, qu'il s'agisse d'un feuillet, d'une double page ou d'un livre. Les liens qui peuvent unir plusieurs modules textuels entre eux sont de natures différentes : il peut s'agir de relations de contiguïté, d'imbrication (lorsqu'un module est intégré, totalement ou partiellement, à un autre). A

cela peut s'ajouter la forme parfois signifiante de la spatialité, comme dans cette nécrologie de Georges-Paul Wagner dans *Présent* qui prend la forme d'un tombeau.

Les relations entre modules textuels peuvent également consister en reformulations polysémiotiques : deux modules, appartenant à des systèmes sémiotiques différents (généralement linguistique et iconique) se font écho l'un à l'autre. C'est le cas de la nécrologie de Robert Parienté dans *l'Équipe*, illustrée par une photographie qui représente Robert Parienté aux côtés de l'athlète Michel Bernard, une complicité évidente liant les deux hommes (la légende indiquant que « Robert Parienté bénéficiait d'une grande complicité avec tout ce que l'athlétisme mondial compte de champions »), qui fait écho à divers passages du texte figurant à côté : « [il était] très proche de tous les hommes et femmes qui firent l'athlétisme français », « [il aimait] les Jeux et surtout ses champions, qu'il célébrait comme ses amis, comme des artistes, en ayant avec eux une relation affective qui en touchera plus d'un ». Ces reformulations polysémiotiques constituent en quelque sorte des renvois non explicités, qui invitent le lecteur sans le dire à mettre en relation les différents modules textuels, du seul fait de leur contiguïté et du regard similaire qu'ils portent sur un même référent.

Ces liens entre modules textuels peuvent également être d'ordre sémantique : dans les textes tabulaires, tous les modules portent sur un même objet de discours, et la cohésion interne est donc très forte, marquée notamment par des isotopies et des phénomènes de reprises anaphoriques. En voici une illustration rapide avec la nécrologie tabulaire de Raymond Devos dans *l'Humanité*, reproduite ci-dessous. Le titre qui surplombe l'hyperstructure dans son ensemble comporte un pronom non encore attribué : « On ne le prendra plus en flagrant délire ». L'objet de l'hyperstructure dans son ensemble sera précisément de donner corps cataphoriquement à ce référent. Soit par l'image : les deux illustrations donnent respectivement un visage et un corps à ce pronom. Soit par le langage, puisque ce « le » est le sujet du discours, à double titre : à la fois celui dont on parle, et le véritable sujet grammatical de la plupart des phrases. C'est ce que montre la titraille, avec par exemple le surtitre du premier article « L'humoriste Raymond Devos est

22 CULTURE

On ne le prendra plus en flagrant délire…

DÉCÈS - L'humoriste Raymond Devos est mort, hier à Paris. Il avait quatre-vingt-trois ans.

CULTURE 23

Confessions d'un artiste

Extraits choisis à l'occasion de longues conversions entre Raymond Devos et notre ami Guy Sèba.

Premiers hommages

mort, hier à Paris. Il avait quatre-vingt-trois ans. », et les intertitres, dont deux sur trois sont des désignations du personnage : « un homme du Nord », « le polisseur de mots ». Ces désignations se font souvent écho, y compris entre les différents modules : ainsi le terme qui désigne le disparu dans le titre du second article, « un artiste », est repris par les derniers mots de l'article du bas de page : « Salut l'artiste ! ». On constate ainsi des phénomènes de reprises anaphoriques et d'échos très forts, qui confèrent une cohésion sémantique à l'ensemble.

Ces divers procédés, qui tissent la textualité de façon originale, peuvent être inhérents au texte (tels les renvois explicites) ou actualisés par l'acte de lecture : c'est en effet souvent le lecteur qui fait exister le texte en tant qu'objet unifié, en activant les relations structurelles, sémantiques ou énonciatives entre les différents modules qui composent le texte.

3. Application aux corpus

3.1 Pourquoi renouveler l'approche du corpus ?

Ce détour par le texte permet de mieux cerner les défauts actuels des corpus, qui souvent sont adaptés uniquement aux textes de type linéaire. L'interpénétration des concepts de texte

et de corpus fait que la représentation de l'un influe fortement sur la représentation de l'autre. Ainsi, la conception actuelle du texte, fortement marquée, comme je l'ai montré, par le primat du linéaire, influence la conception du corpus, qui ne se conçoit que de façon linéaire, ce qui implique de linéariser, d'aplatir les textes tabulaires, de les faire entrer de force dans un moule qui n'est pas conçu pour eux, et de perdre ainsi une bonne partie de leur essence. L'analyse des textes tabulaires est par conséquent limitée, étant donné qu'ils sont traités par des corpus linéaires : tous les composants du texte sont sur le même plan, les discontinuités textuelles sont gommées, la dimension plurisémiotique est évincée.

Par ailleurs, c'est également la représentation du texte qui est biaisée, de façon rétroactive : en tant qu'analystes, nous avons accès aux textes essentiellement par le biais des corpus. Il est par conséquent essentiel que la mise en corpus soit la plus fidèle possible aux caractéristiques des textes d'origine, ce qui permettra d'avoir de ces textes une représentation la plus exacte possible. Le renouvellement de l'approche des corpus se justifie donc du double point de vue méthodologique (les corpus ne reflétant que partiellement les textes qu'ils doivent permettre d'analyser) et théorique (la conception du texte étant faussée par cette fidélité insuffisante des corpus aux textes).

3.2 Suggestions

Il serait donc particulièrement fructueux de prendre en considération dans la mise en corpus la tension que j'ai mise en évidence entre autonomie et interdépendance des différents modules des textes tabulaires. Ceci pourrait à mon sens être permis par la prise en compte de la mise en page dans les textes tabulaires : on pourrait imaginer un logiciel qui isole et numérote chacun des modules textuels, et les mette en relation avec un système de balisage du texte du corpus qui permettrait de distinguer les différents modules du texte entre eux. La mise en relation des résultats de l'analyse statistique avec la mise en page numérisée pourrait donner des résultats éclairants : on pourrait voir notamment entre quels modules la proximité lexicale est la plus forte, ou quels sont les modules qui

renvoient les uns aux autres, et ce, de façon plus systématique que par une étude qualitative.

Ces aménagements me semblent d'autant plus envisageables techniquement qu'il est généralement admis que c'est l'avènement du numérique qui est à l'origine de l'essor du régime de textualité tabulaire, grâce notamment aux liens hypertexte, qui permettent entre autres de créer un réseau de relations entre les modules textuels relevant d'un même texte, fonctionnement qui a été largement exploité par la suite dans les textes sur papier. On peut donc supposer qu'il serait possible (et même plus facile que pour un corpus « à l'ancienne ») de prendre en compte dans la gestion des corpus numériques les discontinuités entre modules propres aux textes tabulaires.

Cette approche renouvelle finalement, quoiqu'à une échelle différente, la problématique soulevée par Damon Mayaffre sur l'articulation du texte à son intertexte au sein des corpus. Il suggérait ici même il y a quelques années de développer les corpus réflexifs, c'est-à-dire des corpus qui intègreraient aux côtés du texte les textes avec lesquels il entretient des relations intertextuelles :

> Les corpus réflexifs devront être organisés techniquement comme des hypertextes : chaque texte constituant devra être relié aux textes considérés comme parents. [...] A l'instar de la navigation sur Internet, la navigation dans le corpus pourra alors s'effectuer. [Mayaffre 2002]

Les pistes que je propose ici vont dans le même sens, si ce n'est qu'elles portent sur une échelle inférieure : il ne s'agit plus en effet de prendre en compte au sein du corpus les relations intertextuelles, mais des relations « intratextuelles »[8]. La tâche paraît d'ailleurs moins titanesque : un corpus ne peut jamais embrasser la totalité d'un intertexte, par essence infini, alors que l'intratextualité se constitue à l'intérieur des frontières du

8 Cette « intratextualité » étant envisagée ici dans un sens différent de celui que lui conférait Genette : il ne s'agit pas ici des relations qui s'établissent entre les différents textes d'un même auteur, mais entre les différentes composantes d'un même texte (et donc notamment, dans le cas du texte tabulaire, entre les différents modules qui le constituent).

texte et est donc plus accessible au bornage imposé par la mise en corpus.

3.3 Gains attendus

Je ne peux malheureusement suggérer de propositions plus abouties quant aux aménagements à apporter aux outils de traitement de corpus, eu égard à mes compétences fort limitées dans ce domaine. Toutefois, si les quelques pistes que je propose ici trouvent un écho auprès de spécialistes du champ jusqu'à être rendues opératoires, on peut supposer que les gains théoriques et descriptifs seraient multiples.

De tels aménagements des logiciels de traitement de corpus pourraient en premier lieu permettre de vérifier les hypothèses présentées ici quant à la théorie du texte. Cette confirmation n'est probablement possible que par le biais d'une confrontation d'un grand nombre de textes, issus de genres différents, entre eux (ce qu'une analyse qualitative seule ne permet pas de faire), afin de vérifier la pertinence de l'opposition entre les deux régimes de textualité que j'ai décrits ici, et de confirmer, voire de compléter, la description que j'ai proposée du fonctionnement du texte tabulaire. Mais le développement des outils d'analyse statistique permettrait probablement de réinterroger également l'autre régime de textualité, car certains aspects du texte linéaire (notamment le rôle de la disposition spatiale et l'interaction entre corps du texte et péritexte) mériteraient d'être envisagés également sous cet angle-là.

Un second enjeu de taille consisterait à renouveler les capacités descriptives des outils de traitement de corpus : aujourd'hui, l'étude quantitative des corpus est encore relativement cantonnée à la lexicométrie. En effet, même si les objectifs de l'analyse sortent souvent du champ de la lexicologie, c'est essentiellement le lexique qui est analysé. Il me semble que la prise en compte de la mise en page, éventuellement de la typographie, serait un premier pas vers un élargissement du champ de l'étude statistique de corpus, qui pourrait être très profitable à l'essor de la linguistique de corpus, en permettant de l'ouvrir aux composantes non exclusivement linguistiques du texte. Je pense ici notamment à la polysémioticité, indissociable du fonctionnement du discours : tout discours comporte

nécessairement une part d'hétérogénéité sémiotique, que ce soit par le biais des mimiques, gestes ou intonations qui accompagnent les mots à l'oral, ou grâce aux images, à la mise en page, à la typographie à l'écrit. Les corpus oraux tendent de plus en plus à prendre en compte cette dimension polysémiotique, mais les corpus écrits continuent de ne s'intéresser qu'au linguistique. Il me semble qu'il y a là un enjeu important pour le développement des corpus, ce qui permettrait par ailleurs d'alimenter un axe de recherche particulièrement dynamique ces dernières années[9].

Le dernier gain attendu serait d'ordre plus pragmatique : le succès des textes tabulaires va croissant et son utilisation fait l'objet de nombreuses recherches, notamment en didactique[10], eu égard à son attractivité et sa lisibilité plus importantes que celles d'un texte traditionnel. Une meilleure connaissance du fonctionnement du texte tabulaire permettrait par exemple de renouveler les dispositifs pédagogiques de l'enseignant (tels que les manuels scolaires) afin de les rendre plus attractifs aux apprenants et de rendre leur assimilation plus efficace.

Ces différents gains attendus (théorique, méthodologique et applicatif) montrent à quel point l'enjeu de la prise en compte de la dimension non-linéaire des textes est important.

---

9 Comme en témoignent les directions de recherche d'équipes telles que le SYLED, dont un des axes du séminaire consiste à s'interroger sur « les problèmes que soulève la plurisémioticité de nombreux discours (on pense en particulier à la presse, aux sites internets, à la publicité…) ; la façon dont des énoncés s'inscrivent dans l'espace (espaces simples comme la page d'un livre ordinaire, mais aussi espaces plus fragmentés du journal, ou du site internet) ; la variation des contenus en fonction des supports d'inscription des messages ; la typographie au sens large de tout ce qui assure une fonction informative de façon non linéaire, etc. » ou Ci-dit, qui organise un colloque en juin 2009 sur le thème *Discours rapporté, citation et pratiques sémiotiques* (« nous proposons d'articuler des approches déjà balisées [...] en renouvelant l'approche du phénomène linguistique par une dimension intersémiotique : en effet si des études existent sur les façons de citer d'autres matérialités que textuelles, elles ont rarement été confrontées de façon systématique aux approches linguistiques et discursives. »).

10 Cf. par exemple Chanier T. & Pothier M. (éds) (1996). *Études de linguistique appliquée* 110 : « Hypermédia et apprentissage des langues ». Paris : Didier.

Conclusion

Au terme de cette étude, j'espère être parvenue à montrer la nécessité de renouveler l'approche traditionnellement linéaire du texte, en dissociant les deux régimes de textualité, linéaire et tabulaire, que j'ai distingués ici. La transposition aux corpus de la description du fonctionnement des textes tabulaires que j'ai esquissée ici devrait permettre une meilleure adéquation des corpus aux textes. Cette adéquation, aussi étroite que possible, entre textes et corpus est fondamentale, dans la mesure où la fiabilité des résultats d'une analyse dépend de la fidélité du corpus aux textes qu'il comprend. C'est en cela que la prise en compte de la mise en page lors de la mise en corpus me paraît intéressante, étant donné qu'elle devrait permettre de mieux analyser les relations originales qui se tissent entre les différents modules d'un texte tabulaire.

Les quelques pistes que j'ai avancées posent toutefois à cet égard des problèmes théoriques et méthodologiques complexes : les diverses discontinuités que j'ai évoquées dans cet article (notamment la polysémioticité, la mise en page) ne se laissent pas analyser aussi aisément que les éléments linéaires. Ces composantes non-linéaires du texte (dont le caractère signifiant est décisif et insuffisamment pris en compte, y compris dans le texte linéaire) sont-elles accessibles à une analyse systématique ? Sont-elles de l'ordre de la langue, du discours, à l'interface des deux ? Il me semble qu'il y a là un enjeu fondamental tant pour la linguistique textuelle (dans la modélisation de ces modes de textualité non-linéaires) que pour la linguistique de corpus (dans la formalisation de ces composantes non-linéaires au sein des corpus, en vue d'une exploration systématique).

Références bibliographiques

Adam J.-M. (1991). *Langue et littérature. Analyses pragmatiques et textuelles*. Paris : Hachette.

Adam J.-M. (1997). « Unités rédactionnelles et genres discursifs : cadre général pour une approche de la presse écrite », *Pratiques* 94 : 3-18.

Adam J.-M. (1999). Linguistique textuelle : des genres de discours aux textes. Paris : Nathan.

Adam J.-M. & Lugrin G. (2000). « L'hyperstructure : un mode privilégié de présentation des événements scientifiques », *Carnets du CEDISCOR* 6 : 133-149.

Adam J.-M. & Lugrin G. (2006). « Effacement énonciatif et diffraction co-textuelle de la prise en charge des énoncés dans les hyperstructures journalistiques », *Semen* 22 : 127-144.

Anis J. (1996). « Texte, textualité, traitements automatiques », *Linx* 34-35 : 257-267.

Audet R. (2000). *Des textes à l'œuvre. La lecture du recueil de nouvelles*. Quebec : Nota Bene.

Clément J. (1994). « Fiction interactive et modernité », *Littérature* 96 [en ligne]. Consulté le 30 avril 2009. URL : http://hypermedia.univ-paris8.fr/jean/articles/litterature.html#fn22

Détrie C., Siblot P. & Verine B. (éd.). (2001). *Termes et concepts pour l'analyse du discours : une approche praxématique*. Paris : Champion.

Genette G. (1987). *Seuils*. Paris : Seuil.

Lane P. (2006). « Pour une reconception linguistique du paratexte », *in* P. Lané (éd.) *Des discours aux textes : modèles et analyses*. Publications des universités de Rouen et du Havre, 183-205.

Laufer R. & Scavetta D. (1992). *Texte, hypertexte, hypermédia*. Paris : PUF.

Lugrin G. (2000). « Le mélange des genres dans l'hyperstructure », *Semen* 13 : 65-96.

Mayaffre D. (2002). « Les corpus réflexifs : entre architextualité et hypertextualité », *Corpus* 1 [En ligne]. Consulté le 30 avril 2009.
URL : http://corpus.revues.org/index11.html

McLuhan M. (1967). La galaxie Gutenberg face à l'ère électronique. Paris : Marne.

Maingueneau D. (1996). *Les termes clés de l'analyse du discours*. Paris : Seuil.

Moirand S. (2006). « De l'aire de la page à l'hyperstructure et à l'écran : comment lire et analyser la presse quotidienne française », *CAUCE, Revista international de Filología y su Didáctica* 29 : 295-320.

Mouillaud M. & Têtu J.-F. (1989). *Le journal quotidien.* Lyon : PUL.

Peytard J. (1975). « Lecture(s) d'une "aire scripturale" : la page de journal », *Langue française* 28 : 39-59.

Piotrowski D. (2004). *L'hypertextualité ou la pratique formelle du sens*. Paris : Champion.

Slatka D. (1985). « Grammaire de texte : synonymie et paraphrase », *in* C. Fuchs (éd.) *Aspects de l'ambiguïté et de la paraphrase dans les langues naturelles*. Berne : Peter Lang, 123-140.

Vandendorpe C. (1999). *Du papyrus à l'hypertexte : essai sur les mutations du texte et de la lecture*. Paris : La Découverte.