

# Exploration de corpus multimodaux pour l'analyse d'un processus de médiatisation : l'exemple du « classement de Shanghai » dans la presse francophone (2003-2010) et de son incidence sur la présentation de soi des universités sur leurs pages Web

Christine Barats<sup>1</sup>, Jean-Marc Leblanc<sup>2</sup>

<sup>1</sup>Université Paris Descartes, Céditec - christine.barats@parisdescartes.fr

<sup>2</sup>Université Paris Est Créteil (UPEC) - Céditec, jean-marc.leblanc@u-pec.fr

## Abstract

This paper intend to articulate a semiological and textometric approach of websites combining text-image dimensions of some home pages of french universities as part of a discursive research on the "Shanghai" mediation process.

## Résumé

Cette contribution entend articuler une approche sémiologique du web et une approche textométrique en combinant les dimensions textes-image de quelques pages d'accueil d'universités françaises dans le cadre d'une recherche discursive portant sur le processus de médiatisation dit de "Shanghai".

**Mots-clés :** Textométrie- multimédia- universités-texte-image.

## Introduction<sup>1</sup>

Cette contribution s'inscrit dans le cadre d'une recherche portant sur l'analyse discursive d'un processus de médiatisation, celui du « classement » dit de « Shanghai ». Elaboré, depuis 2003, par l'université Jiao Tong de Shanghai, ce classement mondial des universités (Academic Ranking of World Universities, ARWU) est un classement académique, issu du champ académique. Il est publié chaque année sous la forme d'une liste de 500 universités. Dans un contexte de médiatisation croissante de classements et de palmarès, ce classement s'est rapidement imposé en France. Nous avons montré l'importance des modalités de constitution de corpus pour étudier un processus de médiatisation et l'imposition d'une formule, telle que la définit Alice Krieg-Planque, la formule : « classement de Shanghai » (Krieg-Planque, 2009 : 63).

Pour ce faire, nous avons interrogé plusieurs bases de données (Factiva, Lexis Nexis) afin de tenir compte des logiques d'archivage qui ne correspondent pas forcément aux logiques du chercheur et constitué un corpus que nous souhaitons représentatif de la presse francophone. Nous avons interrogé ces bases en utilisant différents mots clefs (« classement mondial », « classement international » et « classement de Shanghai ») afin de tendre vers une forme d'exhaustivité de nos corpus et afin de pouvoir étudier la diffusion, la circulation et la stabilisation de la formule.

Dans une approche combinant les résultats de quelques logiciels reconnus dans le domaine tels que *TXM*, *Lexico*, *Tropes Alceste*, *Hyperbase* nous avons formulé des propositions

---

<sup>1</sup> Toutes les données produites ici sont disponibles à l'adresse: <http://textopol.org/contribjadt2012.php>

méthodologiques visant à exploiter ces données au moyen de visualisations et fonctionnalités originales<sup>2</sup>.

En contrepoint de l'analyse de ce corpus textuel de 300 000 occurrences, nous avons examiné les pages d'accueil de quelques universités françaises et étrangères afin de mesurer l'importance accordée à ces classements dans le champ académique, dans le prolongement des travaux menés par Christine Barats. Nous proposons un modèle d'analyse et de visualisation semi-automatisés, reposant sur des méthodes statistiques multidimensionnelles. Nous suggérons en effet que la configuration de ces pages d'accueil, en particulier la présentation de soi (signature, logo, menus, organisation de la page) sont liées à la médiatisation de ces classements qui induisent un souci de soi et une mise en concurrence des établissements. Ces configurations sont révélatrices d'une certaine mise en scène institutionnelle des universités où l'attractivité, la séduction et la logique de marque dominent.

### **1. Présentation numérique de soi, séduction, persuasion, légitimation : rhétorique de la page d'accueil**

Le site qui peut être considéré comme un projet éditorial et un contrat de communication présente en soi une structure complexe. Nous avons appréhendé la page d'accueil, à la fois comme forme ouverte annonçant littéralement (et étymologiquement) une « manière de recevoir » le visiteur, mais aussi comme une composition « techno-sémiotique » dont nous avons examiné les traits. Nous avons privilégié deux axes : la rhétorique de l'accueil et les formes de présentation de soi.

Les pages d'accueil des sites universitaires constituent en effet un matériel d'investigation privilégié pour examiner la façon dont les écrans sont investis et mis en scène pour présenter et identifier l'institution. Elles constituent à la fois une « vitrine », une présentation de soi, et un cadre à l'échange social que constitue l'accueil. C'est pourquoi, nous avons retenu les 17 sites des universités d'Ile-de-France<sup>3</sup> afin d'examiner la façon dont ces institutions académiques se mettent en scène sur le web et mettent en scène leur nom. L'accent portera sur la scène d'énonciation (Maingueneau, 2000 : 69). L'homogénéité de notre corpus relève de l'ancrage territorial, l'Ile-de-France, ainsi que du statut de ces établissements, à savoir leur appartenance à l'enseignement supérieur public<sup>4</sup>. Afin d'adopter une démarche comparatiste, nous avons élargi notre corpus aux sites des quatre universités créées en 1991. Ces universités se distinguent cependant par leur taille et leurs caractéristiques (formations proposées, effectifs et profils socio-démographiques des étudiants, ancrage territorial, recherche, etc.).

La page d'accueil propose une mise en scène de l'énonciation (composition techno-sémiotique et marquage de la relation interpersonnelle). Dans un contexte d'internationalisation éducative et de concurrence numérique, la rhétorique propre à cette valeur de l'accueil contribue ainsi à forger des identités académiques spécifiques dont nous avons étudié les traits techno-sémiotiques distinctifs et les formes de standardisation. L'accent portera sur la scène d'énonciation et la mise en scène de soi (Maingueneau, 2000 : 69). Nous

---

<sup>2</sup> Nous présentons les résultats de cette démarche dans un ouvrage collectif dirigé par Christine Barats pour le Céditec, *Observer et analyser le web en Sciences Humaines et Sociales*.

<sup>3</sup> Nous avons retenu les 13 universités nées après les réformes de 1968 ainsi que les quatre universités créées en 1991 (Université de Cergy-Pontoise, Université d'Evry-Val-d'Essonne, Université Paris-Est Marne-la-Vallée et Université de Versailles Saint-Quentin-en-Yvelines).

<sup>4</sup> Depuis 2006, le dépôt légal a été étendu aux sites web. La BNF propose ainsi un accès aux archives de l'Internet où sont consultables de nombreux sites du champ académique. Les observations présentées dans cet article portent sur la période de septembre 2009 à septembre 2011.

montrons que les signes en présence (images, illustrations, mise en page) répondent à une logique de différenciation dans un contexte de concurrence et « d'économie » dite « de l'attention ». Ces signes appartiennent à différents registres qui mettent en scène l'institution en construisant une forme de communication numérique où la présentation de soi (Amossy, 1999) domine. Le registre de l'histoire (université de Paris), le registre de la science et du savoir (laboratoires, amphithéâtres), le registre de la certification (label) et ceux de l'actualité et du web (réseaux sociaux) constituent les principaux éléments de la configuration techno-sémiotique propre à chaque site. La séduction est au cœur de cette présentation de soi que constitue l'accueil. Les sites ne peuvent plus fonctionner en autarcie, car ils sont inscrits dans une lecture potentiellement simultanée et internationale, ce qui prive toute organisation d'une originalité décontextualisée de l'information qu'elle propose. Ils donnent une matérialité à la notion de concurrence numérique qui rentre en résonance avec la professionnalisation de la communication académique et les mutations qui touchent à l'enseignement supérieur (Loi dite LRU, regroupements d'établissements : PRES, Plan Campus).

### **Quelques éléments d'analyse:**

Corpus et période étudiée (Recherche menée de septembre 2007 à septembre 2011, focus pour l'étude sémiologique sur la période de septembre 2009 à septembre 2011) <sup>5</sup> :

Université Paris 1, Panthéon Sorbonne ; Université Panthéon-Assas (Paris 2) ; Université Sorbonne nouvelle, Paris 3, Université Paris-Sorbonne, Paris IV ; Université Paris Descartes (Paris 5) ; Université Pierre et Marie Curie (Paris 6, UPMC) ; Université Paris Diderot (Paris 7) ; Université Paris 8, Vincennes, Saint Denis ; Université Paris Dauphine (Paris 9) ; Université Paris Ouest Nanterre - la Défense (Paris 10) ; Université Paris-Sud 11 ; Université Paris-Est Créteil Val de Marne (Paris 12), Université Paris-Nord 13, Université d'Evry Val d'Essonne, Université de Cergy-Pontoise, Université de Versailles Saint-Quentin-en-Yvelines (UVSQ), Université Paris-Est Marne-La-vallée.

#### ***1.1. Impératif de communication et formes d'adresse : émergence de standards***

L'émergence de standards s'observe dans la composition des écrans, et en particulier d'adoption du format plein écran qui induit une immédiateté dans l'accueil. Si la rhétorique de l'accueil des sites universitaires se caractérise par un accueil par profils (12 sites sur les 17 étudiés), la pluralité des possibles en termes graphiques et iconiques conduit à une diversité de choix

#### ***1.2. Logo, noms, devises, images : quelles présentations de soi***

Plusieurs signes (signes du prestige académique, devises, label) contribuent à construire des identités académiques distinctives : identité académique héritière de l'historique et prestigieuse *Université de Paris* (Paris 1, 2, 4), identité académique ancrée dans un champ disciplinaire ou un domaine (droit, Paris 2, science, Paris 5, 6, commerce, Paris 9...).

#### ***1.3. Identités académiques et mise en scène de soi : prestige, label, web 2.0.***

Dans un contexte de concurrence numérique, la présentation de soi de la page d'accueil se doit d'être attractive. Le prestige est mis en scène. Des vidéos ou des photos proposent des représentations des lieux et de la vie étudiante. Elles contribuent à rendre attractives ces institutions en proposant une mise en scène de la vie académique idéale : les bâtiments sont accueillants, les chiffres valorisants.

---

<sup>5</sup> Le site de Paris 4 a changé en septembre 2010, celui de Paris 8 en décembre 2010.

Ainsi, l'étude des présentations de soi sur les pages d'accueil des sites académiques indique la prise en compte de logiques de communication et de management, jusqu'alors absentes du champ académique.

## 2. Analyse textuelle des pages d'accueil des universités<sup>6</sup>

### 2.1 Mondes lexicaux, classes sémantico-thématiques des pages web

Nous avons extrait des pages d'accueil des universités le texte de présentation que nous soumettons à une rapide expérience textométrique. Ce corpus est disponible à l'adresse suivante: <http://textopol.org/contribjadt2012.php>. Nous en fournissons ici le protocole de recueil. Les pages de MIT, Oxford et Cambridge sont traduites en français.

SOURCE : site officiel de 12 universités de la région parisienne ; racine ou premier niveau : page d'accueil ou de présentation ou de bienvenue ; texte avec données chiffrées

DATE : 10 octobre 2011

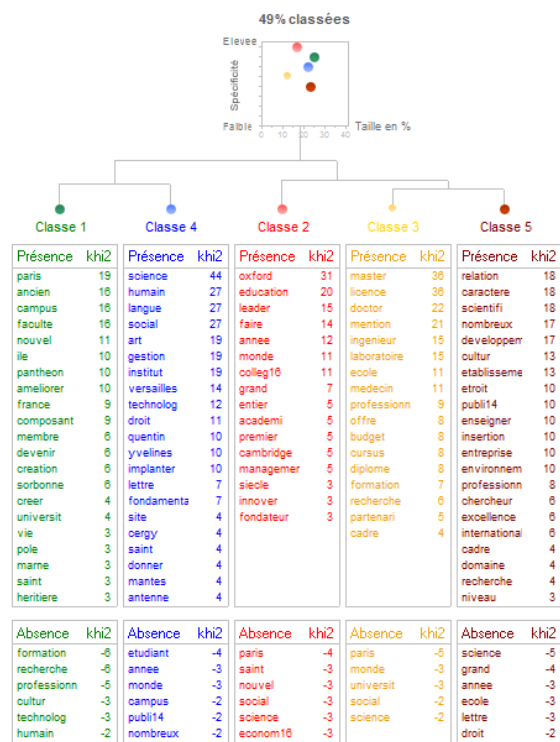
RECUEIL : manuel

FILTRAGE : Critères non strictement formels ; le choix tient compte de l'architecture des sites et vise une comparabilité ; Si la présentation n'est pas synthétique, il ne dépasse pas 2 niveaux de liens dans le recueil.

MISE EN FORME : suppression des liens ; sans illustration ni mise en page ; 19 partitions ; suppression de la signature du président.

CORPUS : 20 sites : 13 uni parisiennes ; une grande école : HEC ; 4 université de la RP non numérotées ; 2 université anglaises

Un premier traitement, réalisé au moyen d'Alceste permet d'identifier quelques mondes



lexicaux. Les résultats obtenus sont bien sûr à considérer avec précautions. Le corpus est très particulier (listes, menus,...) et restreint sur un plan quantitatif. (Selon Lexico 3, 6630 occurrences au total, et une moyenne de 300 occurrences par site).

Nous sommes renvoyés à cette particularité du corpus textuel puisque 49% seulement des énoncés sont retenus par Alceste. La **classe 5** met l'accent sur les *relations*, *nombreuses* que les établissements universitaires entretiennent avec leur *environnement régional*, *social* et *économique* et base la notion d'*excellence scientifique* et pédagogique sur une ouverture vers le monde (Excellence: les réformes sont bien passées par là...). C'est donc ici avant tout le lien vers l'entreprise et la

<sup>6</sup> Une expérience d'aspiration de sites web sur Tropes complètera cette première analyse.

professionnalisation qui est mis en avant. Les Universités Paris 8, Paris 1 et UMPVL sont plus particulièrement constitutives de cette classe.

**La classe 3**, est plus descriptive et porte essentiellement sur l'offre de formation, sans négliger pour autant la notion de partenariat, décidément fortement mobilisée dans le discours universitaire. **La classe 2**, très marquée par la contribution d'Oxford (Premier mot étoilé mais aussi première forme contributive de la classe) mais aussi Cambridge et dans une moindre mesure HEC se caractérise donc par une forte propension à l'autodénomination mais aussi par la présence de deux marques personnelles (*nous* mais aussi *vous*). L'ethos des établissements se construit sur la notion de rayonnement avec les formes *leader*, *premier*, *monde*, *grand*....

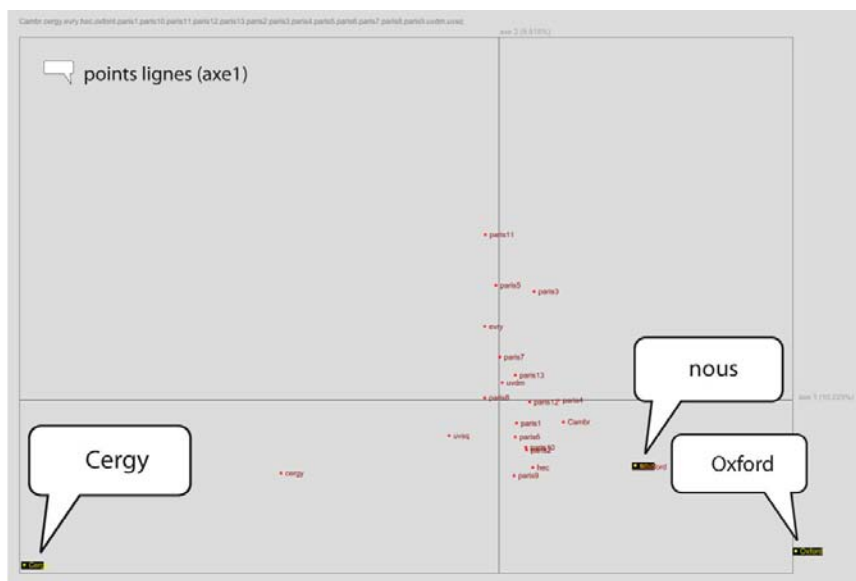
**La classe 4**, majoritairement construite autour des énoncés de Cergy et UVSQ regroupe les énoncés évoquant la place des sciences humaines et sociales mais renvoie également à l'historique et à l'implantation des différents sites et composantes de ces établissements.

**Enfin, la classe 1** est essentiellement construite autour des énoncés des sites de P13 et P12. La forme *Paris*, la plus significative de la classe renvoie là encore à l'auto-désignation des établissements (*université de Paris*, *Paris Sud*, *Paris Est*, *Paris 12 val de marne*...). les formes *PRES*, *pôle* ou *Campus* renvoient aussi à la réforme et au classement.

Il ne s'agit ici que d'un survol de quelques caractéristiques de ces pages d'accueil et une étude complète devra y être consacrée. C'est la démarche que nous souhaitons ici mettre en avant et le projet qui est le nôtre de relier, au sein d'une même étude texte et image.

## 2.2 Approche textométrique des textes des pages d'accueil (Lexico 3/TextObserver)

Sur la base d'une partition en universités, l'analyse factorielle des correspondantes produite par *TextObserver* sur la base d'un tableau lexical produit par *Lexico3* oppose plus particulièrement les sites de Cergy et d'Oxford. La particularité d'Oxford, entrevue au moyen d'Alceste se confirme donc ici et repose notamment sur les formes spécifiques (*Nous*, *grande* et *Oxford*) identifiées par *Lexico 3*. En complément de cette analyse, *TextObserver* nous permet d'afficher les points les plus contributifs de ce premier axe: Les trois premiers: *Cergy*, *Oxford*, *nous*, confirment nos premières remarques. (Liste complète des points contributifs des deux premiers axes en annexe à l'adresse <http://textopol.org/contribjadt2012.php>)



AFC portant sur le texte, partition univ- Axes 1 et 2 (10%, 9%)- *TextObserver*

### 3. Analyse des pages d'accueil: l'expérimentation WebObserver

JP Benzécri a appliqué aux textes des mesures statistiques utilisées depuis longtemps en sciences sociales, comme les analyses factorielles des correspondances, analyses en composantes principales, coefficients de corrélations... Nous proposons de suivre un chemin plus expérimental en empruntant à la textométrie ses principes statistiques multidimensionnels pour décrire les pages web des universités. Le principe de la textométrie, désormais bien connu, consiste à créer à partir des unités textuelles un tableau lexical sur lequel portent les mesures de ventilation, de fréquences, de spécificités des formes lexicales ou d'autres unités mais aussi les analyses multidimensionnelles comme l'analyse factorielle des correspondances.

La maquette *TextObserver*, désormais opérationnelle a pour originalité de proposer des modèles de visualisation originaux comme la présentation tridimensionnelle de l'analyse factorielle des correspondances, l'implémentation du mouvement dans ces représentations, mais aussi la navigation entre différentes dimensions et partitions d'un corpus. Le principe essentiel de *TextObserver* est de considérer d'une manière inédite le tableau lexical, jusqu'ici statique, comme un élément central de l'interface<sup>7</sup>.... La conception ergonomique de *TextObserver* permet de produire très facilement des analyses factorielles sur tout tableau de contingence. Partant de ce principe nous développons une déclinaison de *TextObserver* (*WebObserver*), plus particulièrement dédiée à l'analyse semi-automatisée des sites web. Nous appliquons nos expérimentations sur le corpus des pages d'accueil des sites des universités.

Nous combinons plusieurs niveaux de description en prenant comme point de départ un principe simple: Les pages web que nous voulons analyser comportent des caractéristiques techniques qui sont autant d'éléments que nous pouvons soumettre, sous forme d'un tableau à double entrée, à des analyses factorielles des correspondances ou à d'autres mesures statistiques<sup>8</sup>. Il convient de mener une réflexion sur l'importance ou sur le poids à donner à chacun de ces éléments descriptifs. (Taille de la page en pixels, taille de chaque rubrique, codes hexadécimaux des différentes divisions de la page...). L'outil *WebObserver* est précisément l'outil idéal pour mener ces tests de variation.

#### 3.1. Description par reconnaissance de formes

Nous prenons ici comme point de vue la seule apparence de ces pages en analysant, sous forme d'image ce corpus particulier<sup>9</sup>. Une première expérience consiste à coupler un logiciel de reconnaissance de formes à notre dispositif d'analyse multidimensionnelle. Lors des phases ultérieures du développement de *WebObserver*, il conviendra de développer un module de reconnaissance mieux adapté à la problématique des pages web. Dans cette première phase nous utilisons le logiciel *Roborealm*<sup>10</sup> développé par l'entreprise du même nom, dont nous avons acquis une licence pour les premiers développements de notre module. Ce type de logiciel est développé dans l'industrie et la robotique et est destiné à modéliser l'espace afin de permettre à des robots de se mouvoir dans une zone donnée. Ce type d'analyse induit que le

---

<sup>7</sup> Démonstration et téléchargement sur <http://textopol.org/textobserver>

<sup>8</sup> Une phase importante de cette recherche dont nous exposons ici les toutes premières étapes, consistera à établir si l'AFC est l'outil statistique le plus adapté à ce type de données. (Une expérience par Analyse en Composantes Principales suivra cette première étape).

<sup>9</sup> Le corpus d'images est disponible en ligne à la page précédemment indiquée, rubrique "outils-ressources": visuels universités.

<sup>10</sup> <http://www.roborealm.com>

texte n'est pas considéré ici pour sa valeur linguistique mais bien comme un élément visuel créant un bloc de « couleur typographique », gris visuel résultant de la juxtaposition de noirs, de lettres et de blancs intercalaires.

Nous utilisons ici la fonctionnalité *geometric statistics*<sup>11</sup> qui fournit des données telles que la hauteur et la largeur de la page en pixels, le nombre de pixels autres que blanc entourant une zone considérée ou encore la taille en pixels des différents objets intégrés à l'image. Nous construisons à partir de ces données établies automatiquement un tableau de contingence que nous soumettons à *WebObserver*<sup>12</sup>. Ce tableau comprend 22 colonnes qui correspondent aux 22 sites que nous analysons et autant de lignes que de caractéristiques traitées par le logiciel. Pour éprouver la solidité de notre analyse, nous avons construit une image-test qui permet d'évaluer l'impact des différents paramètres sur la configuration de l'analyse factorielle et qui constitue donc la 23<sup>ème</sup> colonne du tableau.

Le premier axe (31% de l'information totale) démarque clairement le site de Cergy. L'opposition majeure réside entre Cergy d'une part et Oxford/M.I.T/Paris10 d'autre part. Sur l'ensemble du plan on note les groupements suivants, que nous pouvons identifier grâce à la fonctionnalité de zoom de *WebObserver*. Premier groupe: Paris1 / HEC / Paris 8 / Cambridge / Paris 9 / Paris 5 / Paris 11 / Paris 2 / Paris 3 et Marne. Deuxième groupe: UVSQ / Paris 13 / Paris 12 / Evry. Troisième groupe: MIT / Oxford / Paris 10. Enfin Cergy d'une part et Paris 6 et Paris 7 d'autre part. La superposition des images et du plan factoriel avec affichage des points les plus contributifs et explicitation des paramètres utilisés, disponibles en ligne, montrent que l'analyse automatique met essentiellement en évidence des oppositions entre des pages orientées portrait à faible dominante de couleur (Cergy) et des pages à forte dominante de bleu, orientées au format paysage: (Paris 10, MIT, Oxford). Ces données sont encore insuffisantes mais nous suggérons que la démarche mérite d'être explorée et affinée et permettra à moyen terme de dresser de véritables typologies sur lesquelles les sémiologues et spécialistes du web pourront s'appuyer<sup>13</sup>.

### 3.2. Description par *rétro zoning*<sup>14</sup>

#### *Procédures de création*

Lors de la création d'un site web, le zoning est l'une des étapes importantes de la formalisation des recommandations de création du site. Il s'agit d'une étape de décision décrivant l'organisation générale de la page en zones ou espaces d'information caractérisés par rapport à l'espace de la page (résolution en pixels de la hauteur et de la largeur de la future page). Cette étape est primordiale dans la conception d'un projet web sur le plan ergonomique. Le document qui en résulte est volontairement généraliste et peu détaillé.

---

<sup>11</sup> Paramètres complets en ligne sur <http://textopol.org/contribjadt2012.php>

<sup>12</sup> Nous avons dû modifier des données afin de les rendre compatibles au calcul de l'AFC. Le lecteur peut à tout moment vérifier ces données qui sont à disposition sur la page suivante: <http://textopol.org/contribjadt2012.php>

<sup>13</sup> La communication comportera une part de démonstration permettant d'illustrer les fonctionnalités d'interactivité du module *TextObserver/WebObserver* (Zoom, interpolations, variations, affichages ciblés). Une vidéo de démonstration sera par ailleurs mise en ligne à l'adresse citée précédemment. (<http://textopol.org/contribjadt2012>)

<sup>14</sup> Nous devons une grande partie de ce texte à la participation de Marie Pérès, PRAG, UFR SESS à l'université de Paris-Est-Créteil (UPEC), spécialiste de l'image et de la modélisation tridimensionnelle.

L'étape qui suit le zoning est celle de la création de la maquette conceptuelle qui détermine le contenu de chaque zone de manière détaillée. Deux grands types d'orientations sont alors donnés au projet décrit par le zoning de manière très simple : spécifications en termes d'objectif, de comportement (ce que permettra l'interface), spécifications formelles, d'apparence (description de l'apparence que prendra l'élément).

Notre démarche utilisant une procédure à rebours, c'est le résultat vu sur un écran donné (celui de la personne collectant les pages) qui va être analysé. Dans un premier temps il ne sera donc pas tenu compte des subtilités de l'implémentation permettant au visuel de s'adapter ou non à l'écran sur lequel il est vu. Dans un deuxième temps l'exploration (idéalement automatisée) des données de styles du site permettra de déterminer si ces valeurs sont fixes (identiques quelles que soient les dimensions de l'écran ou de l'explorateur utilisé) ou variables (proportionnelles à ces dimensions ou encore variables par paliers).

La procédure de collecte doit donc également faire preuve de rigueur (capture de la totalité de cet écran, sur un navigateur caractérisé, notation de mise en plein écran ou non du navigateur).

Le découpage de la partie utile de l'image résultante doit donc également délimiter automatiquement ce qui appartient à l'explorateur. Un élément à repérer également est la présence ou non d'un ascenseur (« scrolling ») qui modifie la vision de la page et devra être pris en compte lors de l'analyse finale.

Lors de la création d'un zoning<sup>15</sup> plusieurs types de zones sont déterminés. Un premier zoning général est créé qui ne différencie pas les différents types de pages permettant de définir les éléments apparaissant sur tous types de page. Ce zoning est ensuite contextualisé selon la page spécifiée (certaines zones disparaissent ou apparaissent en fonction du type de page).

La page d'accueil fait partie des pages pour lesquelles il est nécessaire d'établir un zoning comportant les zones suivantes :

Identité et positionnement du site, espace de navigation, espaces de mise en avant de contenus, outils, pied de page, espaces publicitaires, auto-promotions.

Nous avons entré dans un tableau les paramètres de ce retro-zoning dont nous fournissons le détail sur le site dédié.

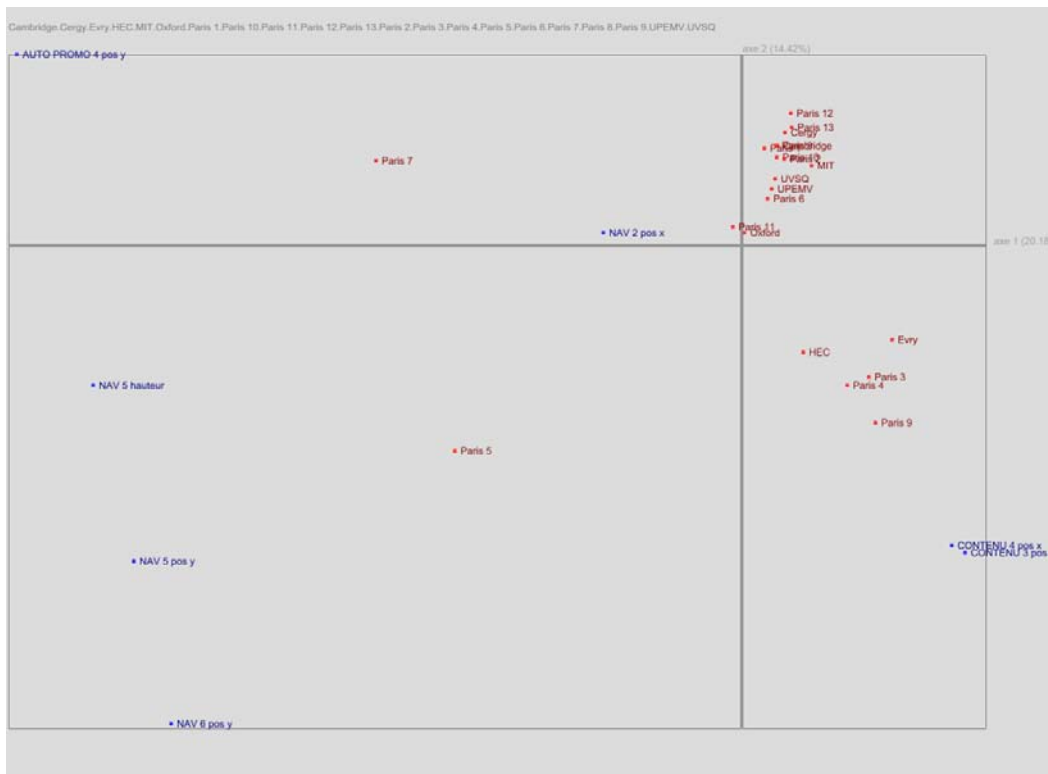
Cette analyse factorielle ne correspond en rien à ce que nous avons produit précédemment. L'analyse des points les plus contributifs affichés sur la figure qui suit explique en partie ce phénomène. En effet, ici ce sont les zones de navigation puis de contenu qui sont discriminantes. Une vidéo en ligne montrera l'incidence de certains paramètres sur la configuration factorielle.

Ici Paris 7 et Paris 5 se distinguent nettement des autres sites en raison du nombre important des champs de navigation.

---

<sup>15</sup> Conjointement à ces zones on peut également déterminer des indications de hiérarchie visuelle. Il s'agit d'un critère ergonomique donnant à une zone un impact visuel plus important que les autres. Cet impact visuel est pris en compte par le directeur artistique lors de la création du visuel. Pour ce qui est de la procédure de rétro-zoning nous ne tiendrons compte de cet élément que dans un deuxième temps en déterminant 4 niveaux d'importances qui peut se traduire en niveaux de gris (0% : Logo, navigation principale; 30% : Corps du document, éléments primordiaux; 50% : Corps du document, éléments informatifs annexes, navigation sur mesure; 80% : Pied de page, navigation générique, outils de recherche).





L'intérêt de *webObserver* est que nous pouvons aisément faire varier plusieurs paramètres de notre tableau de contingence en supprimant une ou plusieurs lignes du tableau, afin de visualiser l'évolution des configurations factorielles obtenues. Il est donc tout à fait envisageable de compiler les trois modes d'analyse dans un même tableau. Cette même conception de *WebObserver* permet également de comparer des AFC produites sur différents tableaux.

#### *Mise en page web et variabilité du visuel*

Les règles ergonomiques d'édition du texte sur le web sont plus des indications que des obligations et visent essentiellement à produire un document qui soit lisible par le plus grand nombre de lecteurs. Les contraintes de lisibilité et de visibilité auxquelles le créateur du site doit chercher à répondre le guideront ainsi vers un choix de famille de police (si possible sans empattement), à une taille particulière (10% supérieure à celle que l'on utiliserait pour une même cible sur un support papier) en privilégiant un fort contraste texte/fond (optimal noir sur fond blanc), et en se préoccupant de la casse (les textes intégralement en majuscule étant moins lisibles ils seront réservés à des titres).

S'interroger sur l'organisation du contenu, sur le rapport texte / image d'une page web nécessite bien sûr de s'interroger sur les positions absolues et relatives de ces éléments ainsi que sur les rapports qu'ils entretiennent. Il est également indispensable de s'interroger sur ce que le concepteur voulait que l'utilisateur voie et sur ce que celui-ci voit réellement. Les deux n'étant pas forcément identiques en raison de la nature mouvante du support page web actualisé sur l'écran de l'utilisateur.

L'une des questions préalables à l'étude d'un affichage écran est celle de la résolution. En effet, si les contenus d'un site web (texte, liens et illustrations) sont entièrement maîtrisés par le concepteur ce n'est pas entièrement le cas pour leur rendu. En effet la configuration de l'écran et des logiciels de navigation définie par l'utilisateur influence l'apparence de la page

web visualisée. Le travail du concepteur doit donc être fait en tenant compte de la variabilité de ces paramètres.

Visiter un site web implique une notion d'action. Le visiteur fait des choix en se déplaçant ou pas dans la page en fonction de sa motivation et du type de lecture qui en découle<sup>16</sup>. Ainsi, il semble nécessaire ici d'interroger l'une des notions importantes en ergonomie web : la notion de ligne de flottaison. Il s'agit de la limite inférieure de la page visible sans faire défiler le contenu à l'aide de l'ascenseur<sup>17</sup>. Les ergonomes web<sup>18</sup> s'accordent à considérer que le contenu principal du site (on doit ici différencier les pages navigantes ou transactionnelles des pages de contenu) doit se trouver au dessus de cette ligne dans la mesure où les internautes ne descendraient en dessous que s'ils y sont invités par des indices leur donnant à penser qu'ils peuvent trouver une information intéressante en dessous de cette limite. Les visiteurs utilisant la page d'accueil uniquement pour naviguer vers un contenu précis doivent donc les trouver au dessus de cette ligne virtuelle, les autres, qui attendent du site qu'il leur fournisse un maximum d'entrées possible vers les pages internes, utiliseront naturellement l'ascenseur à la recherche de clés d'entrée plus ciblées. Les contenus textuels et iconiques du site doivent donc également être interrogés de ce point de vue.

Tout site composé en « portrait » sur un format rappelant la page écrite et dont la largeur est automatiquement calée sur celle de l'écran comportera donc toujours une partie « aveugle ». Les autres possibilités de définition de largeur permettent au créateur de définir la largeur du site soit sur un pourcentage de la largeur de l'écran (dépendant donc de la définition), soit sur une largeur fixe en pixels. Un site composé en « paysage » peut très bien ne pas comporter de ligne de flottaison. Pour cela il est nécessaire qu'il ait été prévu par son concepteur qu'il s'affiche sur un pourcentage de la page (la maîtrise se fait traditionnellement sur la largeur de la page<sup>19</sup> et sera de 100% ou moins) ou qu'il soit en taille fixe et plus petit que l'affichage déterminé par l'écran de l'utilisateur et les choix de ce dernier (ce qui est plus difficile à maîtriser).

Le concepteur peut ainsi déterminer les positions absolues et relatives des éléments iconiques de la page en les disposant dans l'espace. Le texte peut lui aussi être disposé de manière précise mais un certain nombre de problèmes se font jour dès que l'on considère la variabilité du caractère. Un objet textuel dans une page web est soumis à des règles de mise en forme (famille de police, taille, graisse, interlignage, alignement...) qui sont déterminables comme elles le seraient sur un support papier. Cependant il est nécessaire de souligner que, si le concepteur propose la meilleure mise en forme qu'il ait pensé pour le message textuel, c'est l'utilisateur qui dispose dans la mesure où il peut tout à fait supplanter ces choix par les siens en forçant un certain nombre de caractéristiques d'affichage du site web dans son navigateur<sup>20</sup>. De plus l'utilisateur ne possède pas la police prévue par le concepteur celle-ci

---

<sup>16</sup> Le comportement du lecteur varie en fonction de ses préoccupations et de la nature du message. On peut ainsi en distinguer deux types : un comportement de nature qualitative (intérêt plus ou moins régulier pour le message imprimé) et un comportement de nature quantitative (déchiffrement plus ou moins intégral du message qui va d'une lecture intégrale à une lecture de recherche en passant par une lecture sélective) voir schémas du Richaudeau (Richaudeau F., Binisti O., 2005) p77.

<sup>17</sup> Ainsi pour un écran de résolution 1024x768 (la résolution actuellement la plus répandue sur un écran d'ordinateur), cette ligne serait située à 650 pixels du haut de la page.

<sup>18</sup> Dont Amélie BOUCHER (Boucher A., 2007) p.30

<sup>19</sup> Ceci est à relier avec la conception de la page web comme un espace à dérouler verticalement (par contraste avec l'espace du livre à feuilleter horizontalement).

<sup>20</sup> Famille et taille des polices, couleurs, apparence des liens...

sera remplacée par une police de substitution de la même famille si possible. Cette nouvelle police peut très bien faire que le texte occupera l'espace différemment. L'apparence des textes d'une page web est donc éminemment sujette à la variation contrairement à celle d'un support papier<sup>21</sup>. Un bloc de paragraphe peut s'étendre plus ou moins et changer ainsi la composition de la page.

## Conclusion

L'objectif de cette recherche est d'articuler la dimension texte et images, dans une démarche textométrique multimédia et multimodale. Le projet *WebObserver* mobilise des spécialistes de la visualisation, du web de la modélisation tridimensionnelle, de la statistique textuelle. L'analyse des sites web n'est pas ici, comme cela se pratique beaucoup, uniquement orientée sur les liens entrants et sortants ni sur le seul texte constituant les pages mais bien sur l'apparence la mise en page, l'organisation, la structure des dispositifs<sup>22</sup>.

Il s'agit d'articuler une approche sémiologique du web et une approche textométrique. Un des objectifs vise à élaborer des grilles d'analyse pertinentes pour l'étude de corpus web et à moyen terme, de contribuer à la rédaction d'un cahier des charges pour la mise au point d'un outil automatique de description des CMS. Il sera utile de s'interroger sur l'incidence que les normes telles que le W3C (accessibilité) peuvent avoir sur les sites d'institution dans leur mise en page et leur design. Cela crée-t-il des rapprochements ? De la même façon la technologie utilisée et le fait que ces sites sont souvent construits à partir mêmes dispositifs de gestion de contenu (CMS), laisse de moins en moins de place à l'autonomie de concepteur de pages web.

## Bibliographie

- Barats C., « La dénomination comme trace du passé et enjeu : l'exemple des universités parisiennes », revue *Mots*, N°95, mars 2011.
- Barthes R. (1967), « Proust et les noms », *Le degré zéro de l'écriture*, Paris, Seuil.
- Barthes R. (1985), *L'aventure sémiologique*, « *Eléments de sémiologie* », Paris, Editions du Seuil.
- Benzecri, J.-P. (1992), *Correspondence Analysis Handbook*, New-York : Dekker, P. 1994, *L'analyse factorielle*, Paris : PUF
- Boucher A. (2007) *Ergonomie web*. Eyrolles.
- Cibois, P., (1994), *L'analyse factorielle*, Paris : PUF
- Desrosières A. (2008), *Gouverner par les nombres, L'argument statistique II*, Presses de l'Ecole des Mines, Paris.
- Goody J. (1979), *La raison graphique*, Paris, Minuit.
- Granget L. (2004), « Tic et identité de l'université et de l'enseignant : les limites de la communication institutionnelle », *Colloque TICE MÉDITERRANÉE*, Nice.
- Granget L; (2009), *Les universités en quête de prestige dans le grand jeu de la concurrence : le rôle de la communication marketing et l'impact des palmarès*, *Communication et Organisation*, N°35.
- Heilbrunn B. (2007), *La marque*, Paris, PUF (Que sais-je ?).
- Honoré J.-P., Paveau M.-A., Périès G. (2000), éd., « Noms propres », *Mots. Les langages du politique*, no 63.

---

<sup>21</sup> Les documents pdf font relativement exception à cette règle. C'est d'ailleurs l'un des supports recommandés si l'on désire créer un document écran à la mise en page fixe.

<sup>22</sup> Nous pouvons mentionner ici les travaux de G. Fouettilou et de l'initiative Linkfluence. (<http://linkfluence.net>) Linkfluence a développé des méthodes d'échantillonnage des espaces sociaux du web en s'appuyant sur les propriétés structurelles de l'hypertexte.

- Jeanneret Y., Le Marec J., Souchier E. (2003), Lire, écrire, récrire, Objets, signes et pratiques des médias informatisés, BPI.
- Julie Bouchard (sous la direction de), « La communication nombre », revue MEI, Médiation et Information, l'Harmattan, N°28, 2008.
- Kapferer J.-N., Thoenig J.-C. (1994), La marque, moteurs de la compétitivité des entreprises et de la croissance économique, McGraw-Hill, Paris.
- Krieg-Planque A. (2009), La notion de « formule » en analyse de discours, Cadre théorique et méthodologique, Presses Universitaires de Franche-Comté, Besançon.
- Leblanc J.-M., Pérès M. (2010), "Visualiser les données textuelles", TRANSEO, Numéro 02-03 - Mai 2010, URL: <http://www.transeo-review.eu/Visualiser-les-donnees-textuelles.html>
- Maingueneau D. (2000), Analyser les textes de communication, Nathan.
- Miege B. (2007), La société conquise par la communication, t. III, Les Tic entre innovation technique et ancrage social, Grenoble, Presses universitaires de Grenoble.
- Ogien A. (1995), L'esprit gestionnaire, Paris, Editions de l'EHESS, 1995.
- Pailliat I. (1995), L'espace public et l'emprise de la communication, Grenoble Ellug.
- Paveau M. A., Reboul-Touré S., Lecolle M. (2009), éd., « Le nom propre en discours », Carnets du Cédiscor, n° 11.
- Pérès M. (2001). Réflexions sur le modèle informatique du Circus Maximus. Thèse de doctorat, Histoire, Langue et littérature anciennes, Université Michel de Montaigne Bordeaux3.
- Richaudeau F. et Binisti O. (2005), Manuel de typographie et de mise en page, du papier à l'écran, Retz.
- Ruth Amossy (dir), Images de soi dans le discours, la construction de l'ethos, Lausanne, Delachaux et Niestlé, 1999.