

# Web, texte, conversation et redocumentarisation

Jean-Michel Salaün

EBSI – Université de Montréal – CP6128, succ Centre-ville  
Montréal, QC, Canada H3C 3J7

## Abstract

Search engines mainly use linguistic and statistic tools and implicitly regard the Web like a large, global and permanent talk. The Web disturbs the traditional order of documents like a new media taking its place between old ones. A process of redocumentarization is currently taking place.

Usually even with new tools, linguists work within the old world of documents. However they have a special responsibility in building the new one.

**Keywords:** Web, text, document, redocumentarization.

## Résumé

Les moteurs de recherche utilisent principalement des outils linguistiques et statistiques et considèrent implicitement la toile comme le vaste texte d'une conversation mondiale et ininterrompue. Le Web remet en cause l'ordre documentaire comme un nouveau média s'installant sans ménagement parmi les anciens. Un processus de redocumentarisation est en cours.

Les linguistes dans leurs travaux s'appuient généralement sur l'ordre documentaire ancien, même avec des outils nouveaux. Ils ont pourtant une responsabilité particulière pour définir le nouveau.

**Mots-clés :** Web, texte, document, redocumentarisation.

## 1. Introduction

Le texte ci-dessous présente la trame de la conférence. Celle-ci sera développée et illustrée par des exemples, des schémas et des références. Elle s'appuie très largement sur le travail collectif réalisé sous la signature Roger T. Pédauque (2006) au sein du Réseau Thématique Pluridisciplinaire 33 sur le document du CNRS, maintenant en sommeil.

La navigation sur le Web des internautes est aujourd'hui structurée principalement par la dynamique des moteurs de recherche. On peut s'en persuader facilement par les résultats des enquêtes sur les pratiques ou encore par les résultats financiers des firmes dominant le Web, et tout particulièrement de *Google*.

## 2. Texte et conversation

Or les moteurs utilisent principalement des outils linguistiques et statistiques et considèrent implicitement la toile comme le vaste texte d'une conversation mondiale et ininterrompue. Tous les genres, toutes les formes de publications ou d'échanges sont confondus dans un traitement uniforme qui marie une analyse de la langue et un calcul des flux. Ainsi, une requête sur un moteur peut ramener indifféremment un fragment de livre ou simplement sa notice catalographique, un article de journal, un billet ou même un simple commentaire de

blogue, une réponse archivée sur une liste, une rubrique d'encyclopédie, un argumentaire commercial, ou encore bien d'autres genres de textes.

La requête, elle-même, est formulée le plus souvent en texte libre, sans hiérarchie, confirmant la posture conversationnelle entre l'internaute et un vaste robot piochant dans des textes.

Mieux ou pire, le numérique permet aussi un mariage et une production à une échelle inédite entre les images, le son et l'écrit. Cet entrelacement a été préparé par la montée de l'audiovisuel tout au long du XXe siècle, mais la convergence des trois supports de représentation explose littéralement sous nos yeux avec une transformation radicale du statut de la photo, de la vidéo et de la musique, soutenue par les appareils numériques portables et multifonctionnels et aussi par la génération d'internautes nés avec le numérique qui s'en est emparée et s'en sert frénétiquement, comme ses aînés se servaient de l'automobile ou du cinéma dans les années soixante, pour tourner la page de l'ancien monde.

L'uniformisation conversationnelle est encore plus flagrante, si on ajoute que les frontières entre son espace documentaire privé (son « bureau »), la communication dans un groupe, une famille, une collectivité, une organisation, et l'espace public s'effacent de plus en plus. On écrit, on photographie, on publie, on partage, on réagit, tout cela « dans le ciel », c'est-à-dire potentiellement à la vue de tout le monde et on participe ainsi à une vaste conversation mondiale, ou on a l'illusion de le faire.

### **3. Web-média**

Ainsi l'ordre documentaire ancien est écrasé par un média dont la naissance, et la brutale installation entre les modèles traditionnels de la bibliothèque et de la télévision, s'accompagnent à la fois de fulgurances et de confusions pour le sens commun.

Chaque média, bibliothèque, édition de livres, presse, radiotélévision s'est imposé en ébranlant l'ordre économique et la légitimité de ses prédécesseurs. Le Web ne fait pas exception, à ceci près que, cette fois, la rudesse du choc pourrait produire des dégâts plus lourds qu'auparavant. L'uniformisation pointée dans la partie précédente peut aussi, en effet, se comprendre comme le bouclage de cette histoire longue des médias qui s'est brusquement accélérée au siècle dernier.

Le Web, dans sa partie documentaire, emprunte au dernier né - la télévision - bien de ses caractéristiques : un réseau, une vocation universelle dans son audience et son contenu, et surtout son économie marchande : la captation de l'attention de l'utilisateur pour sa revente à des annonceurs intéressés. Mais il s'appuie aussi très largement sur des caractéristiques du plus ancien des médias - la bibliothèque - avec l'accès et la mise en réseau des collections, la distribution particulière des requêtes, entre conformisme et curiosité, rebaptisée aujourd'hui « longue traîne », et surtout le second versant de son économie : la mutualisation des objets documentaires.

Chaque média nouveau a été dénoncé par les gardiens de l'ordre ancien comme dévoyant les valeurs de la culture, du savoir ou de la démocratie. En délégitimant ce qui était légitime, en publicisant ce qui n'aurait pas dû l'être ou qui simplement ne l'était pas dans l'ordre documentaire précédent, le nouveau média perturbe la situation documentaire avant de trouver progressivement sa place dans le paysage général.

On retrouve pour le Web-média les mêmes polémiques qui ont accompagné la naissance de la presse populaire, celles de la radio et de la télévision. Il est un peu tôt pour savoir si celui-là s'assagira à terme, mais on peut déjà repérer facilement à la fois des avancées fortes qui

bousculent une situation sclérosée par des hiérarchies dépassées ou simplement limitée par des outils trop rudimentaires et des dérapages dangereux. Parmi de très nombreux autres exemples possibles, l'évolution de la publication scientifique, jusqu'aux projets de cyberinfrastructures actuels ou encore la place dominante prise par *Wikipédia* chez les étudiants en sont des illustrations.

#### **4. Redocumentarisation**

De leur côté, les chercheurs, qui se penchent sur les discours ou encore la structure de la langue grâce aux facilités de calcul ouvertes par l'informatique, ne s'interrogeaient pas beaucoup sur la légitimité des textes qu'ils étudiaient : celle-ci était donnée par leur statut documentaire et ce dernier était le plus souvent garanti par la forme de l'objet et le traitement qu'il avait subi en amont, ce que l'on peut appeler sa « documentarisation ». Celle-ci s'appuie sur un ordre documentaire issu de l'imprimé et, même s'il ne s'agit là que de sa partie la plus organisée, elle a été systématisée à partir du tournant du XXe siècle par la normalisation de règles de description et de classification, aussi bien du côté des bibliothécaires que de celui des archivistes ou des savoirs administratifs. Pendant tout le siècle dernier et encore aujourd'hui, l'effort s'est poursuivi et élargi avec l'évolution des techniques d'impression.

Ainsi les chercheurs ont réuni pour leurs besoins propres des corpus numériques, souvent d'ailleurs par l'opportunité du moment car ils étaient rares. Ceux-ci représentaient, de fait et implicitement, des collections organisées par une vie documentaire antérieure. Celles-ci, construites donc dans le monde de l'imprimé, ou analogues à ce dernier, impliquaient une quasi-homologie entre texte et document, le premier étant inscrit de façon pérenne sur un support qui lui-même représentait l'objet documentaire.

L'évolution des formats numériques, avec, par exemple, la généralisation du XML qui sépare structure et texte ASCII, autorise des transpositions sur des supports de plus en plus hétérogènes et fragilise la solidité documentaire ancienne en proposant un autre découpage. Le développement des réseaux et l'explosion du Web, qui relie entre eux les fichiers et les textes, qui autorisent des modifications en continu transforment la nature documentaire de ces derniers. Les usages sociaux explosifs et les stratégies économiques lourdes conduisent à remettre en cause les classifications anciennes.

Nous ne sommes plus donc dans cette situation confortable où texte et document étaient presque synonymes. Les conséquences sont sans doute importantes pour les chercheurs qui doivent se poser sur les matériaux qu'ils analysent des questions inédites. Elles sont aussi très lourdes sur l'ensemble de notre relation au savoir enregistré.

C'est pourquoi, il est essentiel de comprendre la dimension documentaire des mouvements en cours. Globalement, nous assistons à une tentative massive de « redocumentarisation » dont il est difficile de prévoir l'issue tant elle est contredite par d'intenses échanges et brassages qui déstabilisent les tentatives. Les appels et les efforts lancés par Tim Berners-Lee pour un Web sémantique, puis pour une science du Web, peuvent être compris dans cette perspective. Mais bien d'autres acteurs et bien d'autres logiques se croisent.

#### **5. Conclusion**

Dans cette présentation, j'ai avancé trois idées complémentaires pour éclairer l'explosion de l'ordre documentaire ancien qui se déroule sous nos yeux et la place des linguistes :

1. Les acteurs dominant du Web le considèrent comme le vaste texte d'une conversation mondiale ininterrompue.
2. Le Web remet en cause l'ordre documentaire comme un nouveau média s'installant sans ménagement parmi les anciens.
3. Une tentative de redocumentarisation est en cours.

Dans ce processus, les linguistes ont une responsabilité particulière, car les outils qu'ils développent avec les informaticiens sont au cœur du mouvement. Peut-être le moment est-il venu pour eux de s'interroger sérieusement sur l'importance de la notion de document pour structurer nos mémoires collectives.

### **Référence**

Pédaque R. T. (2006). *Le document à la lumière du numérique*. C&F éditions.