

### 6.3.2.3. *Le Trésor de la langue française*

Le *Trésor de la langue française* (TLF) est le premier dictionnaire de la langue française à avoir largement utilisé les ressources de l'informatique et cela dès l'origine du projet en 1957, soit bien avant la rédaction elle-même. On connaît le TLF sous sa forme imprimée en 16 volumes publiés en coédition par le CNRS-INaLF à Nancy et Gallimard de 1971 à 1994. Les huit premiers volumes ont encore été composés manuellement, alors que les volumes 9 à 16 ont été photocomposés. Dès 1990, le TLF formate ses données selon la norme SGML. Ce dictionnaire a été élaboré à partir de la base de données textuelles Frantext, qui comme nous l'avons dit plus haut, a procuré aux rédacteurs du TLF plus de 180 millions d'exemples extraits de plus de trois mille textes intégralement saisis en ordinateur depuis plusieurs décennies. Le TLF ne peut donc pas être considéré comme un réservoir de textes à proprement parler, mais plutôt comme un remarquable produit dérivé de la base Frantext. On cherche maintenant à donner au TLF une orientation hypertextuelle, c'est-à-dire de renvois contextuels, notamment en l'enrichissant vers les œuvres contenues dans Frantext. Le TLF sert non seulement de base lexicographique, mais aussi de base de connaissances ; ce dernier volet intéresse particulièrement les chercheurs en intelligence artificielle, notamment l'équipe de Jean-Paul Haton également à Nancy.

Une première version informatisée du TLF a été installée sur un des serveurs de l'université de Nancy en accès public gratuit.

<http://www.ciril.fr/~mastina/> [visité le 5.10.1998]

### 6.3.2.4. *Le projet Athena à Genève*

Le projet Athena est l'œuvre de Pierre Perroud, informaticien et enseignant à Genève. Son site se compose de deux parties : d'une part, d'un ensemble de textes numérisés sous sa responsabilité et, d'autre part, d'une liste importante de références à d'autres textes numérisés.

Les textes numérisés par Athena comportent des œuvres classiques (Corneille, Rousseau, Chateaubriand, Baudelaire, par exemple), ou des textes genevois, comme celui d'Henri Dunant, fondateur de la Croix-Rouge : *Un souvenir de Solférino* (1862), que les spécialistes considèrent comme étant à l'origine du droit humanitaire. On trouve également sur le site Athena des textes originaux universitaires contemporains, ayant généralement trait à l'histoire de Genève et qui n'existent pas sous forme imprimée : Ladislav Mysyrowicz, *Perception des immigrants à Genève vers 1870*.

De façon très classique, les textes sont scannés, puis passés dans un système de reconnaissance de caractères et enfin soigneusement vérifiés. Ils sont ensuite mis à disposition des lecteurs soit en version HTML, soit en version RTF.

Une liste de textes numériques disponibles sur Internet complète le site. Cette liste collationnée par Pierre Perroud est remarquable. Elle renvoie non seulement aux documents numériques créés par Athena, mais également à tous les textes repérés par Pierre Perroud ; elle comprend plus de 3 700 titres. Des liens hypertextes et dirigent automatiquement, à travers les URL, vers les sites et les documents électroniques référencés. Les fonds de nombreux réservoirs de textes

numériques sont répertoriés dans cette liste. Certains sites sont bien connus (Gutenberg, Virginie, ABU, Cornell, etc.), d'autres nous réservent quelques surprises comme le *Naufagé volontaire* d'Alain Bombard en espagnol ou *Tarzan* d'Edgar Rice Burroughs en espéranto ! Elle met aussi en évidence des textes spécifiques à l'histoire de la Suisse (Pacte de 1291 [fondateur de la Confédération helvétique], Constitution fédérale, Chant patriotique genevois du XVII<sup>e</sup> siècle — Cé qu'é lainô —, mémoires de licence en histoire suisse, etc.), ainsi que les auteurs helvétiques (Isabelle de Charrière, Benjamin Constant, Gottfried Keller, Jean-Jacques Rousseau, etc.). Cette liste est donc un outil unique et particulièrement intéressant puisque pour l'instant elle remplace un catalogue collectif qui n'existe pas encore. Nous reviendrons plus loin sur ce point.

<http://un2sg4.unige.ch/athena/html/athome.html> [visité le 5.10.1998]

#### 6.3.2.5. La bibliothèque électronique de Lisieux

Les réservoirs de textes numériques ne sont aujourd'hui plus l'apanage des bibliothèques scientifiques. La bibliothèque électronique de Lisieux a été créée par la bibliothèque municipale de cette ville de Normandie. C'est un exemple particulièrement intéressant et original. Elle n'a pas pour souci d'imiter les importants réservoirs que nous venons de voir, mais cherche à offrir un corpus original de textes appartenant au domaine public. Le patrimoine littéraire et documentaire normand est privilégié. L'ensemble des textes numérisés, disponibles sous la forme de fichiers Ascii, est organisé de la façon suivante :

##### — La sélection mensuelle

Cette rubrique met en valeur un des derniers textes numérisés. De toute évidence, les bibliothécaires cherchent à faire redécouvrir des textes aujourd'hui quelque peu oubliés. En août 1997, il s'agissait de *La Dernière Lettre du matelot* de Georges Eekhoud (1854-1927).

##### — Les archives

On retrouve dans les archives les textes sélectionnés précédemment : Théophile Gautier, Jean Lorrain, Charles Nodier, Ernest Lavisse, etc. Certains textes volumineux sont disponibles sous forme compressée (fichiers zip) facilement récupérables.

##### — Le rayon littéraire

Il offre des textes classiques ou moins connus : Alphonse Allais, Charles Asselineau, Octave Mirbeau, Jean Richepin, etc., ainsi que des bibliographies littéraires.

##### — Le rayon documentaire

Les textes numérisés que l'on trouve sous cette rubrique correspondent à des opuscules, des brochures, des tirés-à-part de revues extraits des fonds anciens de la bibliothèque. Ces textes allant de 1618 à 1877 sont de petits bijoux d'humour et de fraîcheur : nous renvoyons le lecteur à *L'Éloge burlesque de la seringue* (1757) ou à la *Decouverte du style impudique des courtisannes de Normandie* (1618).