

H I S T O I R E



Textes réunis par  
Martine Cocaud

# Histoire et Informatique

Bases de données, recherche documentaire  
multimédia



les  
PUR  
Presses  
Universitaires  
Rennes

## Édition structurée et indexation hypertextuelle d'actes médiévaux

Jacques ANDRÉ et Hélène RICHY  
*Irisa, projet Opéra*  
*Campus universitaire de Beaulieu*  
*F-35042 Rennes cedex*

### Introduction

Les outils de PAO (Publication Assistée par Ordinateur) sont désormais très répandus, ils ont les qualités requises par l'édition en sciences humaines [1] et rendent des services indéniables. Mais, nous ne sommes pas sûrs que ces outils aient apporté quelque chose de vraiment conceptuellement neuf dans le monde de la recherche<sup>1</sup> et particulièrement en histoire. Finalement, entre taper un manuscrit à la machine et utiliser WORD sous PC, il n'y a pas grande différence, mis à part le confort de la correction. De plus, comme les historiens ne sont pas typographes et comme les éditeurs n'emploient plus de préparateur de la copie ni de correcteurs, la qualité des éditions humanistiques baisse sérieusement. Et peut-être aussi le "rendement" des chercheurs.

---

1. On lira sur ce sujet divers articles, sous la plume notamment de Monique PEYRIERE ou de Roger LAUFER, dans le numéro 146 de la revue *Autrement* : "Machines à écrire", juin 1994.

La notion même d'édition est, par ailleurs, remise en cause : non pas par les seuls ordinateurs et leurs traitements de textes, mais par les bases de données et surtout par les possibilités d'accès à des fichiers par réseaux<sup>2</sup>. On se demande même si l'édition traditionnelle d'une charte aura quelque intérêt demain quand on aura accès à un hypertexte de cette charte à condition, bien sûr, d'avoir une meilleure qualité de lecture sur écran et beaucoup plus de facilités d'accès que les seuls langages d'interrogation des bases de données actuelles.

La notion d'hypertexte [7, 11] promet en effet d'aller beaucoup plus loin et de préparer un travail d'historien avec l'aide réelle d'un ordinateur. Mais, la tâche d'édition est alors séparée de celle de création, car souvent incompatible avec les outils d'hypertextes qui ne donnent pas toujours la vision globale de ce que sera l'œuvre finale. Nous proposons — ici dans ce contexte des sciences historiques et d'édition critique — une autre approche qui consiste à considérer les documents comme devant avoir une certaine finalité "éditoriale" : la mise à disposition du public de ces documents, que ce soit sous forme papier ou sous forme hypertexte. Nous supposons alors que ces documents ont une certaine structure (par exemple un découpage en chapitres, sections, etc.). Mais que cette structure permet aussi de considérer finalement tout document comme un manuscrit : vivant, modifiable, actif, grâce à des éléments hypertextuels qui tiennent à jour tous les liens, provisoires ou non.

Pour valider cette approche, nous avons regardé ce que pourrait être l'édition hypertextuelle d'une charte, plus précisément le *Cartulaire de Geoffroy de Saint-Laurent* [22] en nous intéressant notamment à la méthodologie de création et de tenue à jour des tables d'index de façon dynamique. Nous nous sommes appuyés sur un système de documents structurés, GRIF, que nous allons d'abord décrire.

### GRIF et les documents professionnels

Bien qu'étant un éditeur interactif (comme avec WORD ou FRAMEMAKER, on travaille directement sur écran), GRIF<sup>3</sup> ne s'intéresse pas uniquement à la forme graphique des documents : c'est un système de production de documents for-

---

2. Avec des produits comme www, ftp, etc. Un prochain numéro du *Médiéviste et l'ordinateur* doit être consacré à ce sujet.

3. Sous le nom de Grif, on entend en fait deux choses : d'une part un produit commercialisé (par GRIF S.A., 2, bd Vauban, 78053 St-Quentin-en-Yvelines) et, d'autre part, un prototype toujours en cours d'extensions à l'Inria. Nous ne parlerons désormais que de ce prototype. On trouvera dans [2, 9, 14, 16] diverses descriptions de ce système de manipulation de documents.

tement structurés<sup>4</sup>, dont notamment ceux de la documentation technique et scientifique ; ses concepts de base sont proches de ceux de SGML<sup>5</sup>.

Pour GRIF, un document est d'abord une structure logique qui assemble des éléments tels que titres, chapitres, sections, paragraphes, notes, apparat critique, etc. Ces éléments et leurs possibilités de composition sont définis dans des schémas propres à une classe de documents (par exemple les *Que sais-je ?*) à l'aide d'un formalisme proche de celui des grammaires de Chomsky utilisées en linguistique structurale. À chaque type d'élément défini dans un schéma de structure est associé un ensemble de *règles de présentation* qui définissent l'aspect graphique de ce type d'élément. La structure logique est principalement arborescente (hiérarchique). Mais on y a introduit la notion de lien pour donner aux documents une dimension hypertexte [15].

La notion de structuration permet toute une série de calculs (numérotation des sections, des notes en bas de page et de l'apparat critique, etc.), de traitements linguistiques<sup>6</sup>, de vérification de cohérence, de validité, etc. mais aussi de prévoir le travail coopératif de plusieurs personnes sur un même document faisant de GRIF un véritable outil professionnel.

### Index électroniques

La méthodologie traditionnelle (à la main avec fiches, gommes et crayons) de construction d'index a fait l'objet de diverses publications tant en français (par exemple le célèbre *Tapuscrit* [8] ou [13]) qu'en anglais (le non moins célèbre *Chicago Manual of Style* [6]). Le concept d'index électronique<sup>7</sup> se trouve dans divers produits universitaires ou commerciaux<sup>8</sup>. Mais, en général, les tables d'index n'y sont vues que sous leur aspect graphique (en WORD, par exemple, si on clique sur une entrée de la table d'index, on ne voit pas apparaître le texte concerné).

4. Le principe des documents structurés est de distinguer la description logique d'un document de sa structure physique (graphique). On trouvera dans [3, 4] des approches sur ces concepts.

5. SGML, *Standard Generalized Markup Language*, est une norme (définissant la structure logique de documents) de plus en plus employée dans les milieux éditoriaux ; voir [10, 23].

6. Nous avons introduit dans Grif un vérificateur orthographique [18] et y développons un correcteur typographique s'appuyant sur la structure : on vérifiera par exemple qu'il n'y a pas de point final dans un titre, mais qu'il y en a en fin de paragraphe ou pour le dernier élément d'une liste [17].

7. Il ne faut pas confondre "index électronique" avec "indexation automatique", ce dernier terme faisant allusion aux techniques linguistiques ou d'intelligence artificielle pour la recherche informatique des concepts ou mots-clés d'un document ; pour une synthèse, voir [21].

8. On trouvera dans [19, 20] une bibliographie sur ces produits.

GRIF a une philosophie différente : les index sont de vrais liens hypertextuels mais assujettis à la notion de type, ce qui permet des vérifications de cohérence, de validité, etc. [19, 20].

**En pratique :**

- on distingue la mise en page des tables d'index de leur contenu ;
- les tables d'index sont des éléments du schéma de la structure du document, ce qui permet notamment d'indexer une table d'index ;
- les références peuvent représenter soit un renvoi (à une note, à une figure, à une section, etc.) soit une inclusion d'un élément du document ou d'un autre document (permettant ainsi d'indexer plusieurs volumes) ;
- les tables d'index sont tenues à jour en permanence ;
- la présentation des index (emploi du gras, numérotation vers les pages ou les sections, etc.) est guidée par l'utilisateur et non imposée par le système ;
- etc.

### **Le Cartulaire de Geoffroy de Saint-Laurent**

Ayant cherché à tester ces mécanismes d'indexation, nous nous sommes adressés, en 1990, à Lucie Fossier, archiviste-paléographe à l'Institut de Recherche en Histoire des Textes, qui nous a alors fourni un extrait du *Cartulaire de Saint-Laurent* dont elle était justement en train de préparer l'édition. Mais son travail était en phase finale et nous n'avons donc fait qu'une simulation de ce qui aurait pu, quelques années plus tôt, être d'un grand secours.

Le document en question est un cartulaire privé du XIII<sup>e</sup> siècle qui a été découvert "par hasard" parmi des papiers de l'abbaye de Saint-Magloire de Paris par Marc Bloch, qui a été étudié par Anne Terroine et dont l'édition finale [22] a été assurée par Lucie Fossier à la mort de celle-là. Puisqu'il s'agit d'actes très stéréotypés, les éditeurs n'ont pas jugé utile de les présenter dans leur intégrité mais de ne publier qu'un catalogue d'analyses. Il nous semble qu'aujourd'hui la solution serait de fournir l'intégrité de ces actes sous forme électronique accessible sur quelque réseau (ou par CD-ROM), à la fois sous forme d'images du manuscrit original, sous forme transcrite avec apparat critique et sous forme de résumés, avec évidemment tous les outils de consultation utiles, dont ceux que nous allons présenter dans la section suivante.

### **Indexation électronique du Cartulaire de Geoffroy de Saint-Laurent**

Au vu du manuscrit des résumés d'actes et compte-tenu du fait que nous voulions simplement avoir un document sur lequel expérimenter notre méthode d'indexation et non publier ce cartulaire, il nous est apparu qu'un style de

document très simple pouvait suffire. Nous avons ainsi simplement assimilé la succession des actes à la succession des sections d'un rapport de recherche (voir figure 1). Ce travail est complètement décrit dans [5].

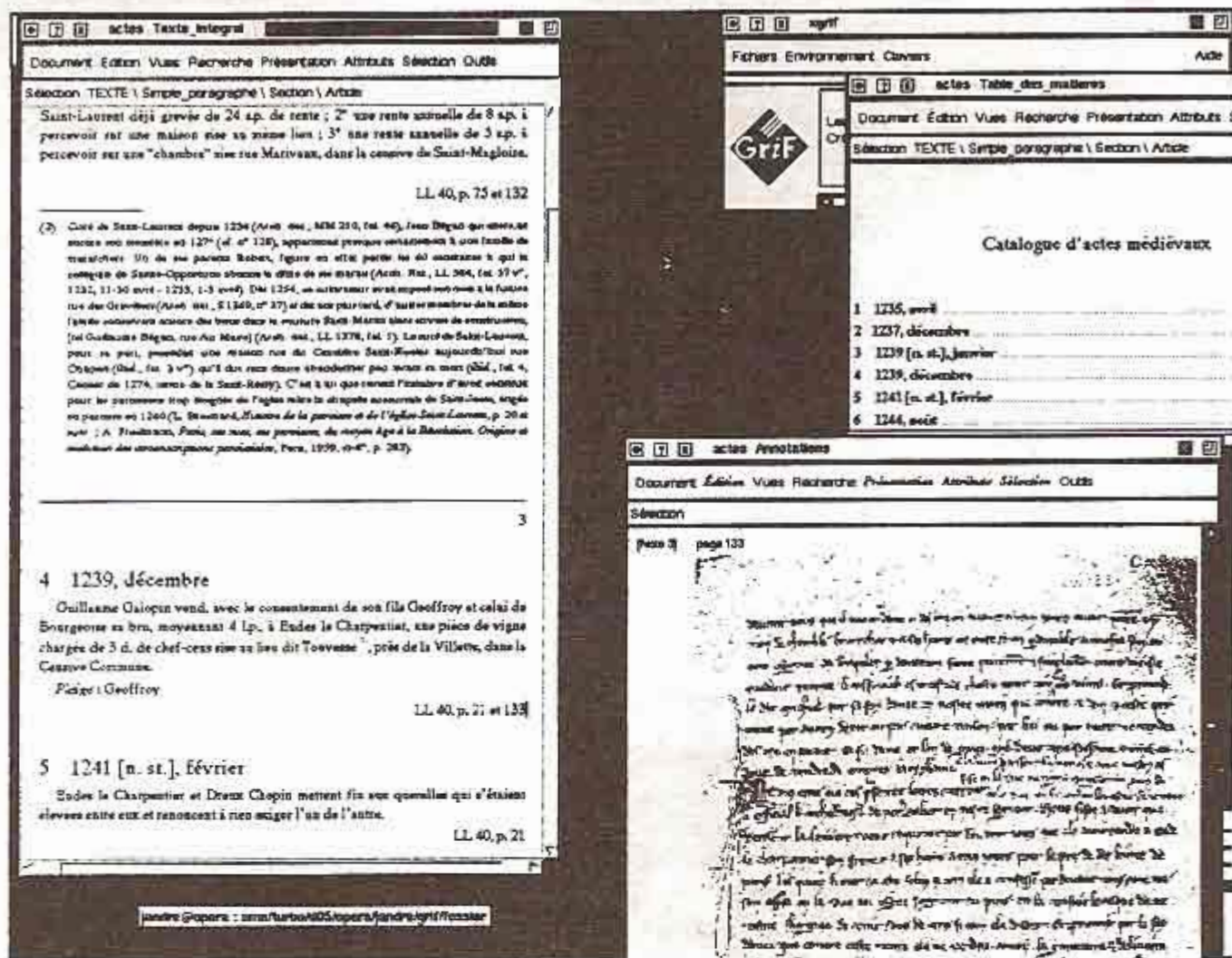


Fig. 1 : Ce que l'on voit sur l'écran (à gauche les résumés d'actes et les notes en bas de page, à droite une image du manuscrit correspondant à l'acte 4).

À tout moment, le lecteur (ou l'auteur en cours de travail) peut visualiser les tables d'index. En figure 2, on voit la table des noms de personnes où on a pris l'entrée "Chrétienne" et cliqué sur "13" ce qui a affiché le résumé de l'acte 13 et inversé (noir au blanc) le mot "Chrétienne" dans ce texte. Dans cette même figure, on voit (entrée "Baudoin Boucel") que l'on peut aussi indexer autre chose que le texte courant, ici une note en bas de page.

Mais c'est surtout en cours de travail que l'auteur peut profiter des facilités offertes : décider des tables (de noms, de lieux, de choses, etc.), de la façon dont elles seront présentées (par exemple noms en italique et renvois aux numé-

ros de pages), de ce qui doit être indexé (la notion d'ancre des hypertextes est remplacée ici par celle de marques qui sont en fait des "parenthèses" englobant un mot, un paragraphe, un titre, etc. au choix de l'auteur), de mettre des gloses (par exemple figure 2 "Chrétienne femme de Jacques d'Arras"), de faire des renvois (figure 2 : "Bégan (Guillaume) : voir Guillaume Bégan" ; en cliquant sur cette entrée on se retrouve sur le renvoi), faire des sous-classifications (par exemple regrouper à "terroir" les divers terroirs cités dans le cartulaire, comme celui de "Ruel", figure 3, qui seront alors triés par ordre alphabétique de leur nom), etc.

<p>6 Catalogue d'actes médiévaux</p> <p>13 1254, novembre</p> <p>Robert, Ermine sa femme, Marie et Emmeline ses sœurs, tous trois enfants et hoirs de feu Isabelle la Fourrée et de feu Hue Fourmentin son mari, vendent à Jacques d'Arras et à sa femme <b>Christienne</b> pour 72 lp., une maison avec son pourpris grevée de 49 s. 4 d.p. de rente, venue aux vendeurs de la succession de la dite Isabelle, sise en face du cimetière de la Trinité<sup>(7)</sup> et tenant d'une part à Gautier Lenfant et d'autre à Ménessier de Ferrières.</p> <p><i>Peine en cas de retrait</i> : 10 lp.</p> <p><i>Pleiges de la peine</i> : Simon de Château-Thierry, cordonnier, et Guérin de Charonne, bourgeois de Paris.</p> <p><i>Pleiges</i> : Adam d'Ouchie et Fromont des Bonnes du diocèse d'Angers. Si le contrat n'est pas respecté, les vendeurs et leurs pleiges encourront la peine d'excommunication.</p> <p>LL 40, p. 171</p>	<p><b>B</b></p> <p>Bandoin Boucel, bourgeois de Paris : (note 10) s. 19</p> <p>Bandoin de Clichy : s. 9</p> <p>Bégan (Guillaume) : voir Guillaume Bégan</p> <p>Bourgeoise, bru de Guillaume Charpentier : s. 4</p> <p><b>C</b></p> <p>Chrétienne, femme de Jacques d'Arras : s. 13, s. 16</p> <p>Christiane : s. 18</p> <p>Colin de Ferrières, frère d'Agnes : s. 19</p> <p><b>D</b></p> <p>Denis, fils de Hersent la Laitière : s. 8</p> <p>Dreux Chopin : s. 5</p> <p><b>E</b></p> <p>Emmeline, fille de Isabelle La Fourrée et de Hue Fourmentin : s. 13, s. 15</p>
--	--

Fig. 2 : En cliquant sur une entrée de l'index, on fait apparaître l'acte correspondant et la sélection est inversée.

En fait, l'auteur travaille essentiellement sur des tables de descripteurs (figure 3) dont les tables d'index ne sont que des représentations graphiques triées. Les champs de ces descripteurs (clés, glose, sujet, etc.) sont à la discrétion de l'auteur. Un historien pourra, par exemple, y ajouter un champ "date" pour y mettre la date de l'acte concerné. GRIF est un système ouvert : ces descripteurs sont accessibles à l'auteur, mais aussi à des programmes. On peut donc écrire (par exemple en TCL [12]) un programme qui mettra tous les noms propres dans l'index, ou un qui fera automatiquement les entrées multiples ("Jean Bégan, curé de Saint-Laurent" donnant les entrées "Jean Bégan", "Bégan (Jean)", "curé de

Saint-Laurent", "Saint-Laurent (curé de)", etc. On pourra aussi écrire un programme recherchant les homonymes (par exemple les diverses "Agnès") ou les noms proches ("Adam d'Ouchy" et "Adam d'Ouchie") et, compte tenu de critères de dates et de lieux, aider l'historien à regrouper ou ventiler ces divers noms. Mais c'est l'historien qui décide. La notion d'annotation<sup>9</sup> est alors un complément bien agréable pour y noter ses interrogations ou suppositions.

<p>8 1250 [n. st.], janvier  <b>Hersent la Laitière</b> vend à Eudes le Charpentier, son frère, moyennant 10 lp., une maison chargée de 20 a.p. de "cens", sise rue du Cygne et attenant au puits, dans la censive de l'évêque de Thérouanne<sup>9</sup>, la dite vente étant faite avec le consentement de Denis, Hervi et Auberée, fils et fille de la dite Hersent et de Gautier son gendre.          (Official de l'archidiaacre)</p> <p style="text-align: right;">LL 40, p. 135</p>	<table border="1"> <tr> <td>[63]</td> <td>«</td> <td>clé (s) de tri :</td> <td>(1) terroit, (2) de Ruel,</td> </tr> <tr> <td></td> <td></td> <td>sujet :</td> <td>lieu</td> </tr> <tr> <td>[64]</td> <td>«</td> <td>clé (s) de tri :</td> <td>(1) Hersent la Laitière.</td> </tr> <tr> <td></td> <td></td> <td>sujet :</td> <td>personne</td> </tr> <tr> <td></td> <td></td> <td>glose :</td> <td></td> </tr> </table>	[63]	«	clé (s) de tri :	(1) terroit, (2) de Ruel,			sujet :	lieu	[64]	«	clé (s) de tri :	(1) Hersent la Laitière.			sujet :	personne			glose :	
[63]	«	clé (s) de tri :	(1) terroit, (2) de Ruel,																		
		sujet :	lieu																		
[64]	«	clé (s) de tri :	(1) Hersent la Laitière.																		
		sujet :	personne																		
		glose :																			

Fig.3 : Descripteur associé à une entrée en cours d'indexage.

## Conclusion

Le gros avantage d'utiliser un système structuré avec des liens hypertextuels pour indexer un document est que l'on profite de la notion d'hypertexte mais aussi de tous les avantages, cachés, des documents structurés : les références sont continuellement mises à jour et si on ne peut pas référencer n'importe quoi (les références sont donc toujours valides), par contre on peut référencer, au choix, une section, un mot, un chapitre, une page, etc. En même temps, l'édition triée des tables en fait un outil très facile à utiliser.

Mais au-delà des possibilités offertes par un système comme GRIF<sup>10</sup>, ce qui nous paraît le plus important est qu'il considère un document comme actif, c'est-à-dire quelque chose de vivant dont la forme physique (livre, écran...) n'est qu'une image à un instant donné.

9. Il s'agit de commentaires électroniques que l'on peut attacher de façon provisoire à un endroit du document, comme les *post-it* des secrétariats. Cette notion, présente dans GRIF, sera notamment accessible dans les postes de lecture de la Bibliothèque de France [24].

10. Certaines lacunes sont sûrement encore présentes dans ce système et nous sommes donc prêts à faire d'autres expérimentations, en collaboration avec des chercheurs en sciences humaines, pour améliorer cet outil.



## Bibliographie

- [1] Jacques ANDRÉ, "Traitement de texte et histoire des textes", *Le médiéviste et l'ordinateur*, XVI, p. 19-33, automne 1986.
- [2] Jacques ANDRÉ, D. DECOUCHANT, V. QUINT et H. RICHY, "Vers un atelier éditorial pour les documents structurés", *Actes du congrès Afcet 1993*, vol. 4, p. 63-72 (aussi paru en Rapport de recherche Inria, numéro 1971, août 1993).
- [3] Jacques ANDRÉ, Richard FURUTA and Vincent QUINT, *Structured documents*, Cambridge University Press, 1989.
- [4] Jacques ANDRÉ et Vincent QUINT, "Structures et modèles de documents", *Le document électronique*, Cours Inria (sous la direction de Christian Bornes), p. 3-57, juin 1990.
- [5] Jacques ANDRÉ et Hélène RICHY, *Utilisation des index d'un éditeur structuré dans le cadre d'actes médiévaux*, Publication Interne Irisa n° 841, Rennes, juin 1994.
- [6] *The Chicago Manual of Style*, The University of Chicago Press, Chicago, IL, USA, 13th edition 1982.
- [7] Marie-Catherine DANIEL-VATONNE, "Hypertextes : des principes communs et des variations", *TSI — Technique et Science Informatique*, vol. 9, n° 6, 1990, p. 475-492.
- [8] M.-L. DUFOUR, *Le tapuscrit, recommandations pour la présentation et la dactylographie des travaux scientifiques (sciences humaines)*, École des hautes études en Sciences sociales, diffusion CID, 1971.
- [9] R. FURUTA, V. QUINT and J. ANDRÉ, "Interactively Editing Structured Documents", *Electronic Publishing — Origination, Dissemination and Design*, 1 (1), p. 19-44, avril 1988.
- [10] Michel GOOSSENS et Éric van HERWIJNEN, "Introduction à SGML, DSSSL et SPDL", *Cahiers GUTenberg*, (12), p. 37-56, décembre 1991.
- [11] Roger LAUFER et Domenico SCAVETTA, *Texte, hypertexte, hypermedia*, Que Sais-je ? n° 2 629, Presses Universitaires de France, Paris, 1992.
- [12] John K. OUSTERHOUT, "TCL : An Embeddable Command Language", *Winter Usenix Conference Proceedings*, 1990.
- [13] Benoît PATAR, *Directives aux auteurs pour la confection d'un manuscrit*, Le Préambule, Canada, 1990.
- [14] Vincent QUINT and Irène VATTON, "Grif : An Interactive System for Structured Documents", *Text Processing and Document Manipulation*, J.-C. van VLIET, éd., p. 200-213, Cambridge University Press, 1986.
- [15] Vincent QUINT and Irène VATTON, "Combining hypertext and structured documents in Grif", *ECHT'92 : Proceedings of the ACM Conference on Hypertexts*, D. LUCARELLA, J. NANARD, M. NANARD and P. PAOLINI, éd., p. 23-32, ACM-Press, USA, 1992.
- [16] V. QUINT, I. VATTON, J. ANDRÉ et H. RICHY, "Grif et l'édition de documents structurés : nouveaux développements", *Cahiers GUTenberg*, (9), p. 49-65, juillet 1991.

- [17] Hélène RICHY et Jacques ANDRÉ, *Correction typographique et édition électronique*, Publication interne, Irisa, Rennes, 1994 (à paraître).
- [18] Hélène RICHY, Patrice FRISON, and Éric PICHERAL, "Multilingual String-to-String Correction in Grif, a Structured Editor", *EP92*, C. Vanoirbeek and G. Coray, éd., p. 183-198, Cambridge University Press, 1992 – voir aussi : *idem*, "Intégration d'un correcteur orthographique dans l'éditeur structuré Grif", Publication Irisa n° 676, 1992.
- [19] Hélène RICHY, *Grif et les index électroniques*, Rapport de recherche n° 1 756, Inria, 1992.
- [20] Hélène RICHY, "A hypertext electronic index based on the Grif structured document editor", *Electronic Publishing - Origination, Dissemination, and Design*, 7 (1), p. 1-14, march 1994.
- [21] George SALTON, *Automatic Text processing*, Addison-Wesley, reading MA, USA, 1989.
- [22] Anne TERROINE, *Un bourgeois parisien du XIII<sup>e</sup> siècle : Geoffroy de Saint-Laurent, 1245 ?-1290*, CNRS Éditions, Paris, 1992 (édité par Lucie FOSSIER).
- [23] Dominique VIGNAUD, *L'édition structurée de documents — SGML — application à l'édition française*, Éditions du Cercle de la Librairie, Paris, 1989.
- [24] Jacques VIRBEL, "Reading and Managing Texts on the Bibliothèque de France station", in P. Delany and G. Landau (eds.), *Text Based computing in the Humanities*, MIT Press, 1993, p. 75-94. Voir aussi "Écrire pour relire", *Autrement*, n° 146, juin 1994, p. 149-153.