

Segmentation d'articles mathématiques en vue d'une navigation

Boucif, BOUKACEM*, Thierry LAFOUGE¹, Hubert EMPTOZ*

¹URSIDOC : Unité de Recherche en Science de l'Information et Documentation
43, bd du 11 Novembre 1918 - 69 622 Villeurbanne cedex Bât Omega.

*LIRIS : (Laboratoire d'InfoRmatique en Image et Systèmes d'information)
Bâtiment Jules Verne - 403 -
Institut National des Sciences Appliquées de Lyon
69621 - VILLEURBANNE Cedex

1. Introduction

Pour introduire cet article, il est pertinent de citer les propos de BERTRAND-GASTALDY : « L'avènement de l'ordinateur et surtout des micro-ordinateurs a augmenté de façon considérable la capacité des organisations et des individus à générer de l'information contenue dans les documents de manière générale et dans le document mathématique en particulier. Mais si la plupart des documents sont depuis plusieurs années mis en forme avec l'aide de l'ordinateur, leur gestion, depuis leur production jusqu'à leur consultation en passant par l'analyse de leurs structures, leur diffusion et leur stockage est loin de répondre aux besoins réels des utilisateurs [1] ».

1.1 Quelques constats préliminaires

« Les journaux scientifiques ont été longtemps le premier instrument de la diffusion scientifique. En 1995, le passage du modèle papier aux modèles électroniques s'impose, spécialement dans le domaine des Mathématiques et de l'Informatique car l'environnement de travail des deux communautés, celle des informaticiens et, de façon plus récente celle des mathématiciens, intègre l'Internet depuis plus de dix ans. Les publications périodiques électroniques concernent aujourd'hui la totalité des chercheurs et les produits sont suffisamment importants et variés [2] ».

L'écriture des textes mathématiques utilise toujours deux dimensions et l'avènement de l'informatique et du numérique n'a pas résolu tous les problèmes liés à cette spécificité. Certes on ne peut ignorer les progrès liés à l'avènement de LATEX, mais les problèmes liés à la diffusion et la manipulation sur le Web des

documents mathématiques restent complexes. La production de documents scientifiques est liée aux développements d'environnements graphiques permettant d'exprimer et de manipuler une grande diversité d'objets (textes simples, formules mathématiques, schémas); des problèmes particuliers se posent encore pour représenter et afficher les formules mathématiques qui ont une structure complexe et bidimensionnelle (fractions, matrices, intégrales. Le traitement des objets mathématiques soulève de nombreux problèmes tels que la complexité et la diversité des règles typographiques, la gestion des grandes formules, l'ambiguïté des notations mathématiques.

1.2 Enjeux de la segmentation des articles mathématiques

Caractères spécifiques et aspect pérenne des mathématiques

La documentation joue un rôle très important dans tous les domaines scientifiques, ce rôle est primordial en mathématiques. Un article publié dans une revue mathématique contient en effet l'intégralité d'un résultat ; ce dernier sera immédiatement utilisable par la communauté scientifique.

L'utilisation d'articles (et de résultats) anciens, datant parfois de plusieurs décennies (voire plus), est courante en mathématiques. La pérennité des articles mathématiques est une de leur spécificité, cela tient au fait que les mathématiques ne constituent pas une science expérimentale où tout peut être remis en cause par l'expérimentation ; ainsi, les œuvres de Henri Poincaré qui sont toujours d'actualité et ses travaux sur les fondements de la géométrie ("On the Foundations of Geometry ", 1898) ou sur les mathématiques et la logique (Les mathématiques et la logique, 1905 et 1906) [3] sont des outils de travail pour les mathématiciens.

Les mathématiciens pensent que les mathématiques ont un caractère « imprévisible », car certains résultats ont parfois contre toute attente, un impact important sur les sciences, l'ingénierie ou les mathématiques elles-mêmes, des années après leur découverte : c'est la raison pour laquelle ces derniers attachent une importance à la documentation. Les mathématiques constituent une discipline particulière parce qu'elles constituent une science des modèles et une base d'outils pour les sciences physiques et/ou expérimentale.

Cette importance de la documentation est renforcée par l'avènement de l'ère électronique. En effet, de nouvelles possibilités sont offertes, telles que, les bases de données pour l'organisation et les revues électroniques pour la diffusion et l'accès à cette dernière. Comme le souligne F. Bérard : « Notre préoccupation essentielle pour l'avenir des mathématiques doit être de préserver la diversité des sources bien identifiées et fiables ». « L'ensemble de la littérature mathématique, constitue en fait l'environnement dans lequel les mathématiciens vivent et trouvent une source essentielle d'inspiration, de concept et d'outils pour faire aboutir leurs idées [4] ». Ces propos renforcent l'idée que les chercheurs forgent leurs inspirations et leurs intuitions essentiellement dans la documentation, même ancienne. C'est dans le corpus des résultats acquis qu'ils trouvent leurs outils. C'est souvent en faisant des rapprochements inattendus entre champs, a priori éloignés, qu'ils font progresser leur science. La documentation constitue donc pour les mathématiques une ressource « stratégique » à laquelle la communauté consacre beaucoup de soins, de temps et une part importante de son budget de recherche. Ce besoin vital a certainement été une des motivations majeures pour la rédaction du traité intitulé "Nicolas Bourbaki, Eléments de mathématiques" ; cette œuvre devait regrouper l'ensemble de la connaissance mathématique et être une sorte d'ouvrage de références.

Nécessité de segmenter pour l'accès à l'information

Le travail du mathématicien consultant les articles mathématiques relève davantage de la recherche d'informations que de la lecture intégrale ; ils recherchent des informations particulières à des problèmes spécifiques ; leur mode de travail peut consister en la recherche de théorèmes, la lecture des introductions ou des résumés bien ciblés qui leur permettent d'aller consulter directement tel ou tel article, formule ou démonstration ; la mise en forme du "Bourbaki" évoqué précédemment était conçue pour faciliter ce mode de travail avec de plus la présence de nombreux liens et renvois.

Pouvoir accéder en un temps optimal à ces parties de textes que nous appellerons unités autonomes est donc un enjeu important ; l'objet de cet article est de présenter une méthode de segmentation qui permettra de localiser et d'identifier ces parties autonomes

Après avoir retracé brièvement, l'historique et l'importance de la documentation mathématique pour les chercheurs, nous allons fixer les objectifs et la démarche en vue de cette segmentation, nous illustreront ensuite avec deux exemples la méthode de découpage proposée.

Nous nous situons dans le contexte classique de segmentation de documents. Pour cela, nous nous inspirons en partie des travaux de M. BESSONNAT [6], qui a défini le découpage en « paragraphes et ses fonctions », et de ceux de M. OUFELLI [7] qui a utilisé la notion de paragraphe pour découper le document technique. Il a choisi le paragraphe comme unité minimale pour former les différentes unités documentaires répondant aux propriétés requises (autonomie linguistique, cohésion syntaxique et sémantique) formant un bloc « thématique » et homogène.

Notre problématique est alors la suivante : comment faire pour segmenter (découper) un article mathématique ? Que choisir comme unité documentaire ?

Quel niveau de granularité faut-il atteindre pour proposer à l'utilisateur des unités autonomes qui soient à la fois :

- porteurs d'un sens (compréhensibles, mais brèves) ;
- exploitables (donc pas trop longues) ?

Quels sont les critères de découpage qui vont nous aider à faire le bon choix ?

Quels sont les indicateurs nécessaires à un bon découpage ?

Il convient de signaler que la segmentation à laquelle on s'intéresse ici est différente de la segmentation envisagée par les spécialistes du traitement des images et de la rétroconversion ; leur activité de segmentation a pour objet, dans un premier temps, de séparer le document en texte d'une part et formules d'autre part et ensuite à décomposer chacun en caractère élémentaire que l'on reconnaîtra, TOUMIT [5].

2. Notre démarche en vue de la segmentation

2.1 Introduction à la segmentation

Nous avons élaboré la stratégie de segmentation présentée ici en nous appuyant sur une analyse d'usage enrichie par l'analyse d'un questionnaire proposé aux mathématiciens (lequel est présenté en annexe) et en l'expérimentant sur un corpus d'articles mathématiques ; ce corpus est composé de 30 articles, rédigés en Français et provenant des revues Elsevier¹ (Éditions Scientifiques et Médicales Elsevier), qui édite une vingtaine de revues mathématiques, SMF² (Société Mathématique de France) et la SMAI³ (La Société de Mathématiques Appliquées et Industrielles).

Après avoir relevé des indicateurs et marqueurs dans le texte de ces articles, nous avons étudié trois méthodes de segmentation :

- méthode de segmentation en fonction de la structure rédactionnelle,
- méthode de segmentation utilisant les marqueurs de mise en forme (ponctuation),
- méthode de segmentation en fonction des spécificités mathématiques.

Il est clair que ces trois méthodes ne sont pas totalement indépendantes, dans la suite du présent article nous présentons la dernière.

Les mathématiciens ne peuvent connaître la totalité des mathématiques (même d'une partie!), ils doivent naviguer d'un document à l'autre dans la littérature à la recherche de concepts, d'idées, d'exemples, de théorèmes, de résultats, de démonstrations diverses. Plus important encore, ils doivent s'appuyer sur des résultats antérieurs qu'ils ne veulent pas toujours vérifier par eux-mêmes. Les propos extraits d'entretiens avec les mathématiciens lyonnais de l'Institut Girard Desargues ou de l'Ecole Normale Supérieure illustrent bien leur méthode de travail.

« Je ne lis pas un article pour rédiger, mais je me réfère à un article déjà lu. D'où la primauté donnée aux théorèmes et corollaires ».

¹ <http://www.elsevier.fr/>

² <http://smf.emath.fr/>

³ <http://smai.emath.fr/>

« Cela dépend de l'usage que je veux en faire. Par exemple si je cherche un résultat particulier je ne regarde que les théorèmes, propositions et corollaires ».
« Je commence par lire le résumé, puis l'introduction, puis je survole les explications intermédiaires », « Je fais surtout une lecture non linéaire. Ceci dit, je cherche ce que je pense être le point-clé de la démonstration et je l'étudie en détail ».

Ces quelques propos montrent bien que les mathématiciens ne procèdent pas à une lecture intégrale et continue ; ils ont des parcours qui vont d'unités documentaires en unités documentaires, les expressions telles que théorèmes, corollaires, propositions, résumé, introduction, explications intermédiaires, démonstrations constituent l'essentiel de leur discours. Pour utiliser les travaux des autres, le mathématicien cherche à accéder uniquement aux parties les plus pertinentes (ou plus exactement qu'il considère comme telles dans le contexte précis de son travail), à savoir, théorèmes, lemmes, démonstrations,...

2.2 Les spécificités du langage mathématique

Nous allons maintenant définir avec précision ce que nous appelons « étiquette mathématique » dans un article, et ceci en nous appuyant sur l'observation de notre corpus.

L'étiquette mathématique sera représentée par deux items :

- son libellé,
- son énoncé.

Nous avons relevé dans notre corpus, les différents types d'étiquettes mathématiques « dites ici de premier niveau » : « Théorèmes, Lemmes, Remarques, Propositions, Corollaires, Définitions ». Nous définissons également une étiquette de deuxième niveau : « Preuve ou Démonstration ». Cette dernière est toujours accompagnée d'une étiquette de premier niveau.

Nous avons trouvé d'autres étiquettes mathématiques de premier niveau qui ne figurent que dans quelques articles telles que : « Applications, Énoncés, Affirmations, Notations, Hypothèses, Conjectures, Exemples », nous les prenons en considération, lors du découpage. Nous définissons la notion de libellé et d'énoncé.

2.2.1 Le Libellé

Le libellé est une étiquette mathématique, suivi ou non par un numéro.

Nous avons quatre (04) cas différents dans notre corpus :

1^{er} Cas :

Théorème 3.1 ou bien THÉORÈME 3.1

Le mot théorème peut être en minuscule ou en petite majuscule.

{3} : représente la section 3

{1} : peut être soit le premier théorème de la section 3, soit la première étiquette mathématique de la section 3.

2^{ème} Cas :

Théorème 1 ou bien THÉORÈME 1

{1} : représente soit le premier théorème soit la première étiquette mathématique de la section en cours.

3^{ème} Cas :

Théorème ou bien THÉORÈME

Dans ce cas nous n'aurons de numéro ni sur les libellés, ni sur les sections.

4^{ème} Cas :

Théorème A ou THÉORÈME A

Dans ce cas, nous avons des libellés avec des lettres alphabétiques.

2.2.2 L'énoncé

L'énoncé est constitué d'un libellé et de la partie qui le compose. Cette partie qui suit le libellé va l'expliquer. Elle est pour de nombreux cas en italique.

Exemple 1:

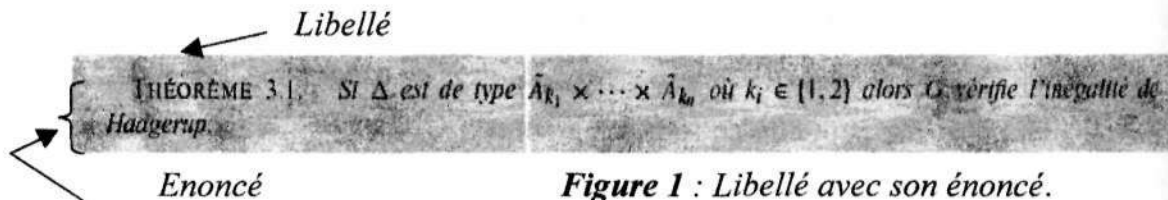


Figure 1 : Libellé avec son énoncé.

Exemple 2 :

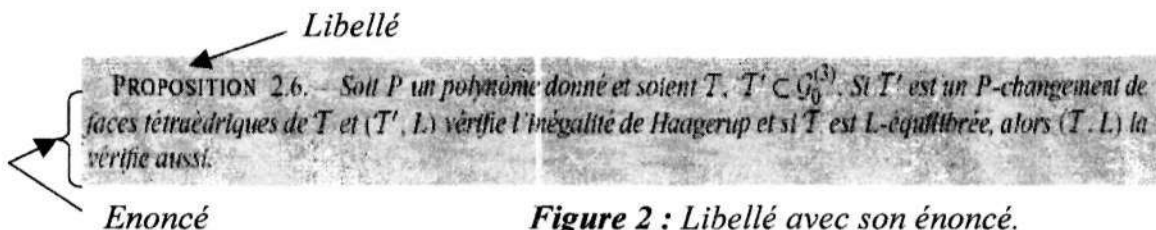


Figure 2 : Libellé avec son énoncé.

2.3 Principe et règles de segmentation de la méthode

Nous définissons tout d'abord ce que nous entendons par segment. Un segment d'un article est une partie de ce dernier, il est délimité par des marqueurs de début et de fin que nous appellerons « balises » qui sont de « nature différente » suivant le type de méthode utilisée. La segmentation consiste à découper l'article en un ensemble de segments disjoints et dont la réunion recouvre l'article tout entier. Les marqueurs devront être précis et sans ambiguïté afin que l'on puisse automatiser le processus de segmentation. Ce dernier s'accompagne en général de recherche de liens avec d'autres segments. Nous faisons l'hypothèse qu'il est préférable de segmenter, puis de rechercher ensuite la présence de liens implicites ou explicites entre les segments.

Nous donnons ici le schéma général de notre démarche en vue de la segmentation des articles mathématiques. Dans cette article, nous ne traitons que la phase 1.

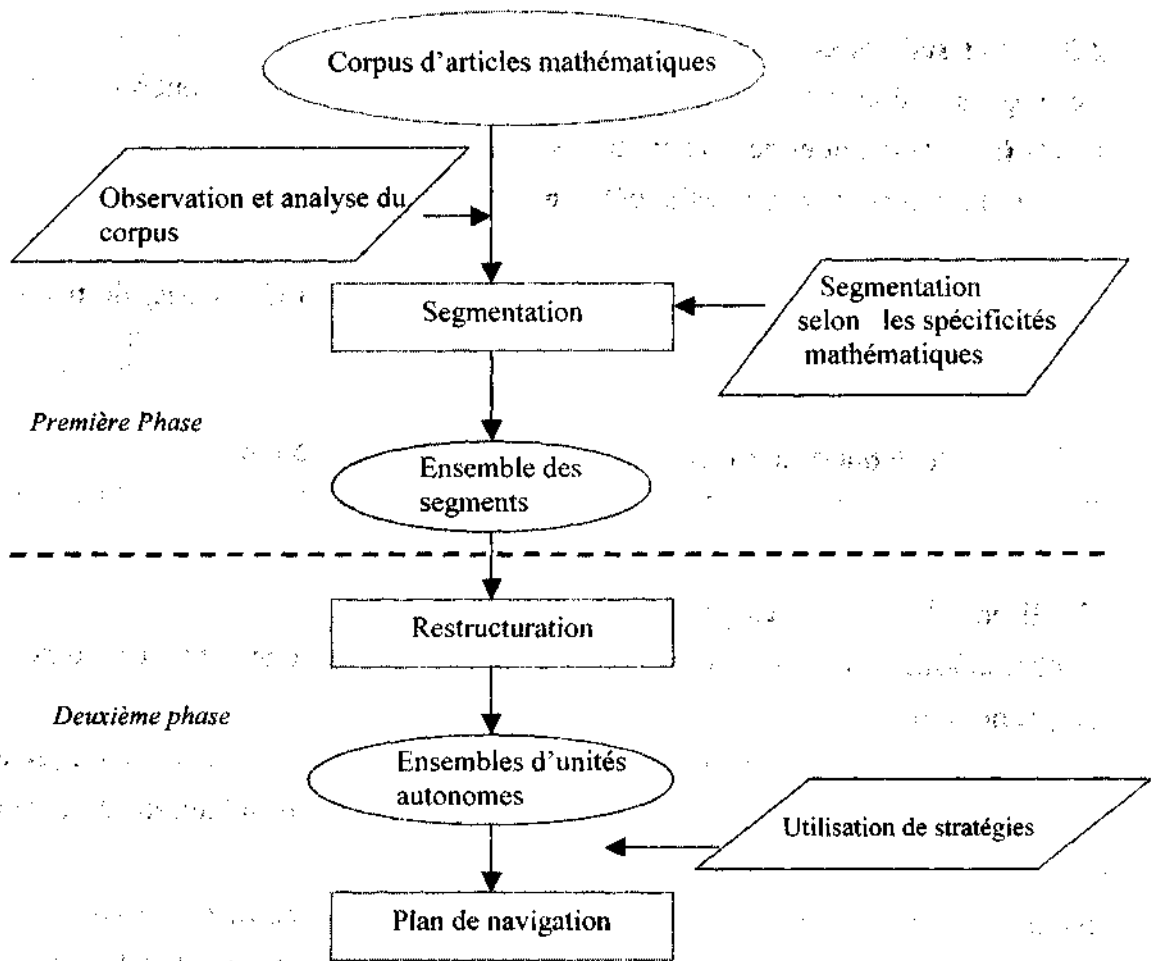


Figure 3 : Schéma général

Nous définissons trois types de segments :

- Les segments étiquettes (de niveau 1) qui se déclinent suivant différents types : théorèmes, lemmes, ...
- Les segments étiquettes (de niveau 2) : preuves ou démonstrations.

Si nous découpons en ne prenant compte que ces segments, nous négligeons un volume important de texte de l'article. Aussi, nous définissons un autre type de segment que nous appelons « segment texte ».

- Les segments texte : ce sont des parties de texte, qui ne peuvent pas être reliées de façon implicite ou explicite aux deux types de segments que nous venons d'introduire.

2.3.1 Les segments étiquettes (de niveau 1)

Les segments étiquettes sont caractérisés par un « libellé + un énoncé ».

Les balises de début et de fin sont telles que :

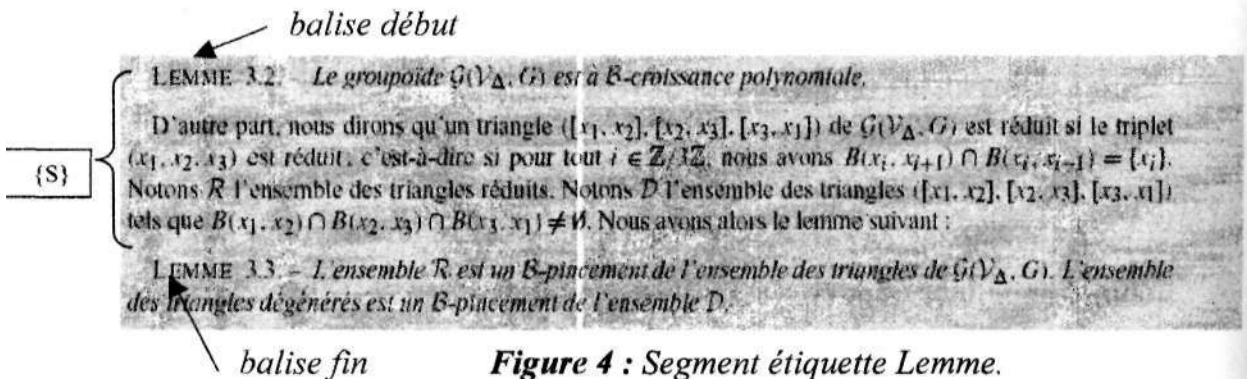
- La balise de début est le début du libellé d'une étiquette mathématique,
- La balise fin est :
 - le début d'un autre libellé d'une étiquette mathématique, de niveau 1 ou 2,
 - ou
 - le début d'une section⁴, sous section ou sous sous-section.

Dans le cas du format L^ATEX [8], qui est très adapté à la rédaction de documents structurés, une raison pour laquelle la plupart des mathématiciens l'utilisent. Il existe des balises indiquant le début et la fin d'une étiquette mathématique et qui permettent d'automatiser facilement ce type de segmentation.

Nous avons relevé dans notre corpus, plusieurs types de segment étiquette : théorèmes, lemmes, corollaires, propositions, résumés, introduction, définitions, applications, notations, remarques, exemples, références.

Nous allons donner des exemples de segment étiquette pour quelques cas.

Exemple d'une étiquette mathématique ayant pour balise de début « le début d'une étiquette mathématique » et pour balise de fin « un autre libellé ».



balise début

{S} LEMME 3.2. - Le groupoïde $G(V_\Delta, G)$ est à \mathcal{B} -croissance polynomiale.

D'autre part, nous dirons qu'un triangle $([x_1, x_2], [x_2, x_3], [x_3, x_1])$ de $G(V_\Delta, G)$ est réduit si le triplet (x_1, x_2, x_3) est réduit, c'est-à-dire si pour tout $i \in \mathbb{Z}/3\mathbb{Z}$, nous avons $B(x_i, x_{i+1}) \cap B(x_i, x_{i-1}) = \{x_i\}$. Notons \mathcal{R} l'ensemble des triangles réduits. Notons \mathcal{D} l'ensemble des triangles $([x_1, x_2], [x_2, x_3], [x_3, x_1])$ tels que $B(x_1, x_2) \cap B(x_2, x_3) \cap B(x_3, x_1) \neq \emptyset$. Nous avons alors le lemme suivant :

balise fin

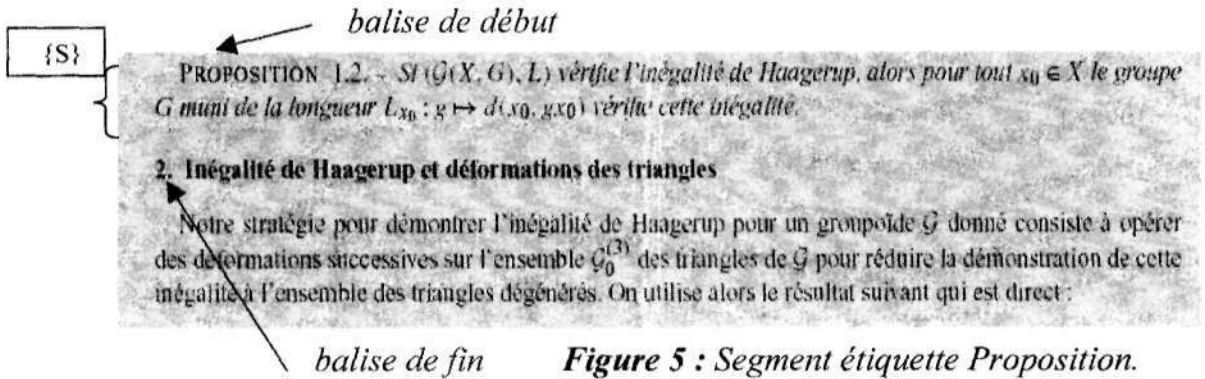
LEMME 3.3. - L'ensemble \mathcal{R} est un \mathcal{B} -piacement de l'ensemble des triangles de $G(V_\Delta, G)$. L'ensemble des triangles dégénérés est un \mathcal{B} -piacement de l'ensemble \mathcal{D} .

Figure 4 : Segment étiquette Lemme.

⁴ La structure logique d'un article mathématique en latex est structurée en sections, sous sections, sous sous-sections, paragraphes, sous paragraphes.

Comme représenté sur la figure 4, la balise de début du segment étiquette {S} est « LEMME 3.2 », et la balise fin est : « LEMME 3.3 ».

Exemple d'une étiquette mathématique ayant pour balise de début « le début d'une étiquette mathématique » et pour balise de fin « le début d'une section ».



Comme représenté sur la figure 4, la balise de début du segment étiquette {S} est « PROPOSITION 1.2 » et la balise de fin est la balise de section « Inégalité de Haagerup et déformations des triangles ».

2.3.2 Les segments étiquettes (de niveau 2)

« Preuves ou Démonstrations » : sont caractérisés par un « libellé + un énoncé », à la différence avec les étiquettes de niveau 1, ils ne possèdent en général pas de numéro.

Les balises de début et de fin sont définies telles que:

- La balise de début sera le début du libellé étiquette mathématique « *Preuve ou Démonstration* ».
- La balise fin est
 - Le marqueur graphique « \square » (celui-ci, est généré par Latex), ou le marqueur lexicale « CQFD » (Ce Qu'il Fallait Démontrer),
 - ou
 - le début d'un autre libellé, d'une section, sous section ou sous sous-section.

Prenons un exemple d'une étiquette mathématique de niveau 2 :

Balise début

Démonstration. – Là encore, elle est immédiate, en se reportant à la formule d'Itô étendue, et au calcul de base :

pour tous processus matriciels adaptés A et B , on a

$$\left\langle \int_0^t \text{tr}_N(A_s dH_N(s)), \int_0^t \text{tr}_N(B_s dH_N(s)) \right\rangle = \frac{1}{N^2} \int_0^t \text{tr}_N(A_s B_s) ds,$$

$$\left\langle \int_0^t \text{tr}_N(A_s dH_N(s)), \int_0^t \text{tr}_N(B_s dK_N(s)) \right\rangle = 0.$$

Remarquons que $\text{tr}_N P(H_N, K_N)$ et $\text{tr}_N Q(H_N, K_N)$ définissent bien des processus réels, car P et Q sont auto-adjoints. \square

Figure 6 : Segment étiquette démonstration.

balise de fin d'une démonstration

Remarque :

Dans notre corpus, nous rencontrons en général les segments « preuves ou démonstrations » à la suite des segments étiquettes de types « théorèmes, lemmes, propositions ». Cependant, les démonstrations ne sont pas toujours contiguës, un marqueur linguistique dans le texte de l'article nous le précise. Ce marqueur linguistique peut se trouver dans le titre d'une section⁵.

Titre de section annonçant la preuve d'un théorème

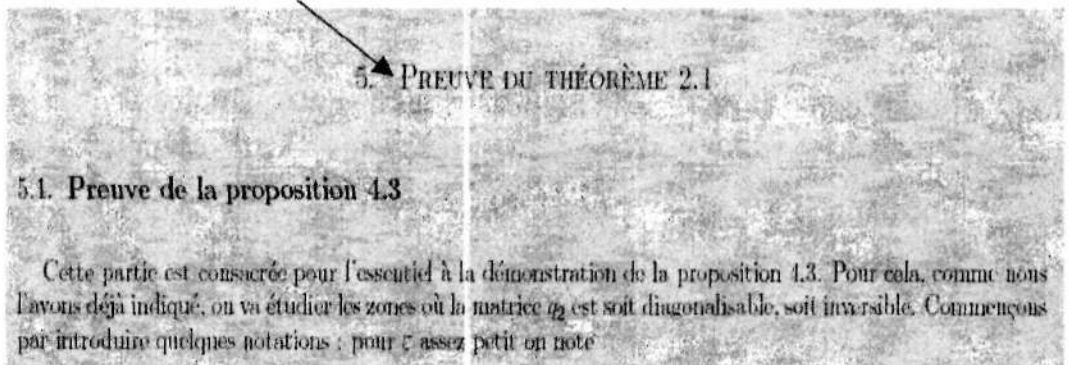


Figure 7 : Segment étiquette Preuve.

Pour cet exemple de la figure 7, extrait du corpus de notre thèse, nous avons un théorème (THEOREME 2.1) qui se trouve dans la section 1 de l'article alors que sa

⁵ La structure logique d'un article mathématique en latex est structurée en sections, sous sections, sous sous-sections, paragraphe, sous paragraphe.

preuve se trouve dans la section 5. Nous pouvons remarquer que nous avons ici une réelle difficulté pour les mathématiciens de parcourir tout l'article mathématique à la recherche de la preuve du théorème. Nous pouvons remarquer dans cet exemple l'intérêt de la méthode de la segmentation. En effet lors de la segmentation nous allons créer une unité autonome qui sera composé du segment théorème 2.1 (étiquette de niveau 1) et du segment étiquette preuve du théorème 2.1, qu'il faudra construire.

2.3.3 Les segments texte

Les segments que nous appelons dans cet article segments texte sont des segments qui sont délimités par deux balises, une balise début et une balise fin, nous avons relevé dans notre corpus deux types :

Premier type de segment texte :

- une balise de début qui peut être soit une section, soit une sous section soit une sous section,
- une balise de fin qui est le début d'un autre libellé d'une étiquette.

Exemple : un segment texte qui a pour balise de début la section «1. Définitions, notations» et pour balise fin le libellé d'une étiquette de premier niveau ici « DEFINITION 1.1».

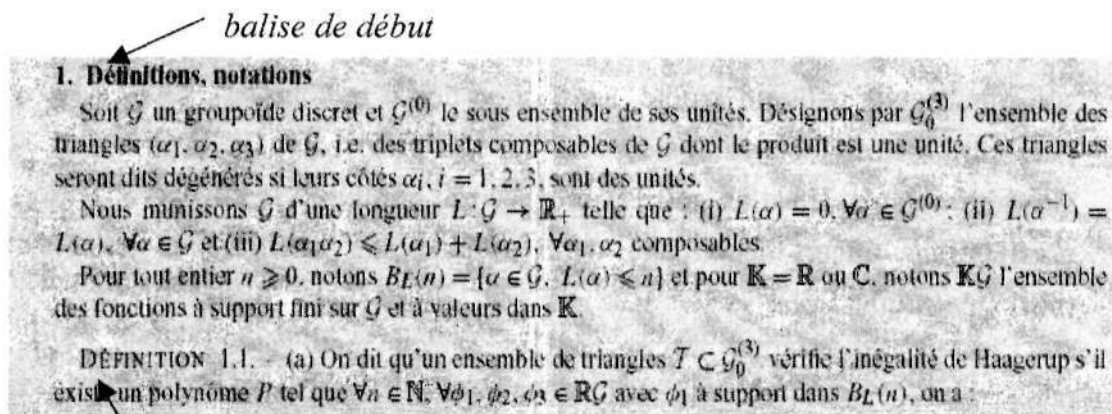


Figure 8 : Segment texte.

balise de fin

Deuxième type de segment texte :

- une balise de début qui est une balise de fin d'un libellé d'une étiquette de second niveau.

- une balise de fin qui est le début d'un autre libellé d'une étiquette,

Exemple d'un segment texte dont le début est la fin d'une étiquette de niveau 2, ici le caractère □, et la fin le début d'un autre étiquette de niveau 1, ici « PROPOSITION 1.8» :

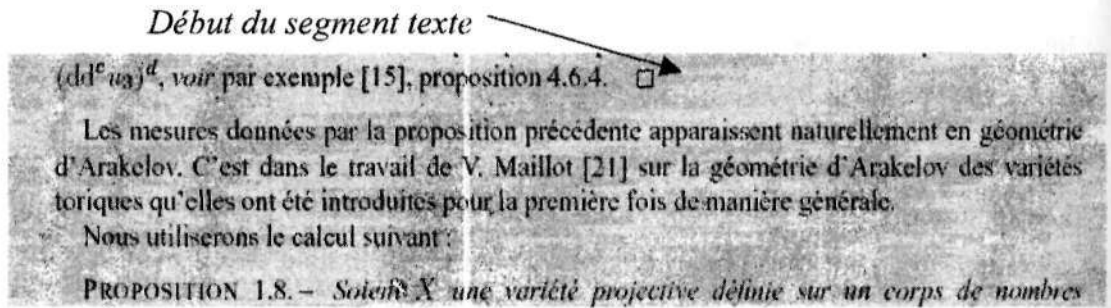


Figure 9 : Segment texte.

2.4 Mise en œuvre de la segmentation selon les spécificités mathématiques

Pour illustrer notre travail de segmentation, nous allons choisir deux exemples que nous segmentons selon cette méthode.

Nous commençons par un cas simple, nous avons choisi un extrait d'article de Mr TALBI (cité dans la bibliographie). Nous commençons notre segmentation en fonction de notre méthode par découper le résumé, comme segment à part, vu son importance dans un article. Le résumé décrit le contenu exact de l'article et permettra de donner au lecteur les idées principales et les résultats qu'on voudrait obtenir. Il est en général bref, ce qui conduit à le conserver entièrement n'effectuons aucun découpage. Nous faisons de même pour les références bibliographiques.

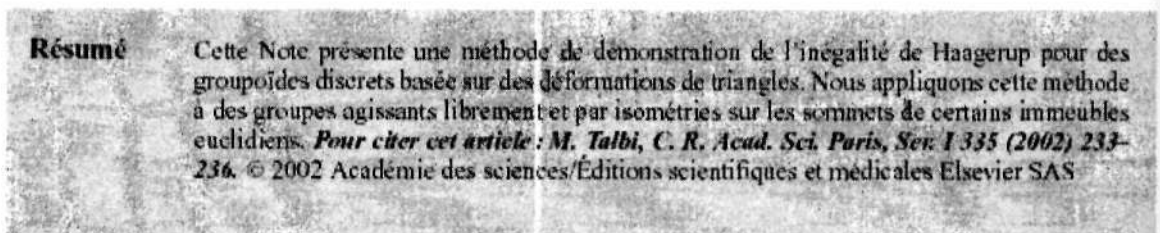


Figure 10 : Résumé.

Ensuite, nous découpons cet extrait d'article en fonction des étiquettes mathématiques, c'est à dire, on parcourt l'article et à chaque rencontre d'un des segments définis auparavant nous découpons.

2. Inégalité de Haagerup et déformations des triangles

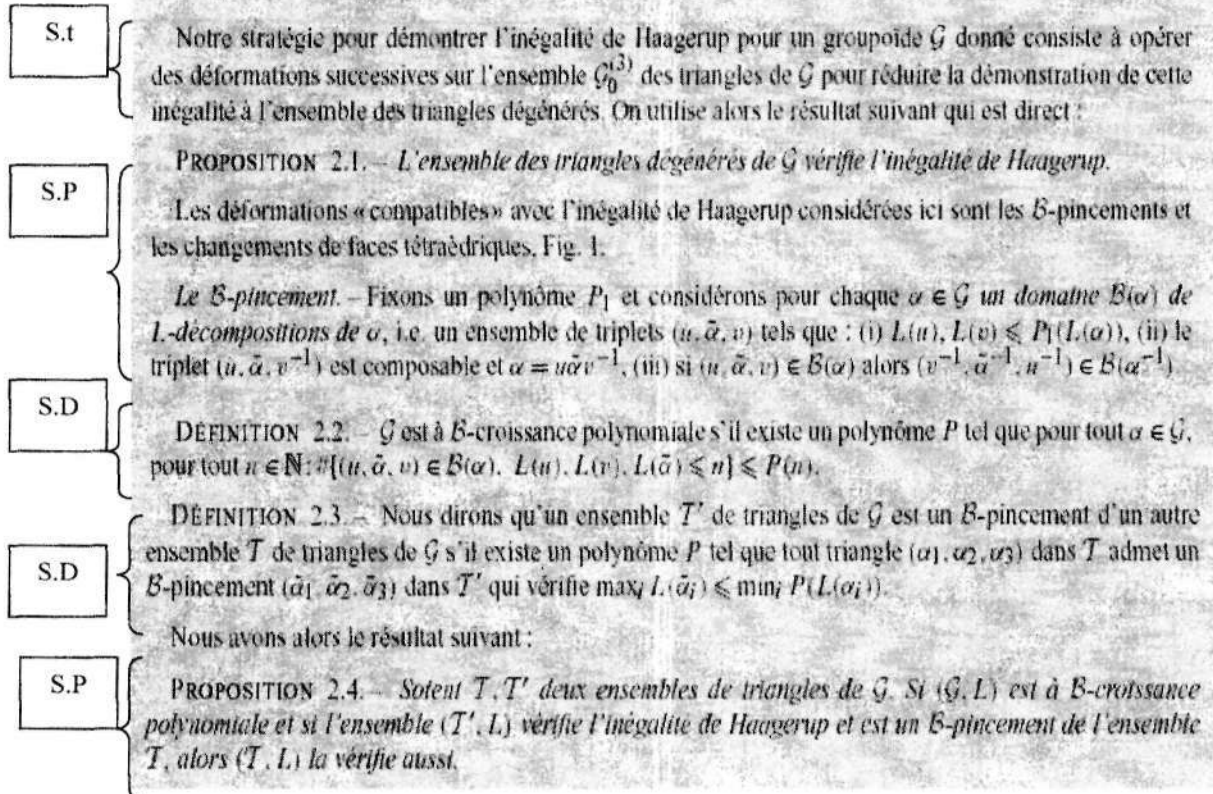


Figure 11 : Exemple de découpage.

Sur cette figure, nous présentons quelques segments ici de notre découpage, St : Segment texte ; SP : Segment Proposition ; SD : Segment Définition.

Nous allons maintenant faire le même travail sur un autre article plus long, plus complexe, qui possède différents niveaux⁶, celui de Mr. BELLASSOUED (cité dans la bibliographie). Nous allons le découper en prenant en compte les trois niveaux, et en le présentant sous forme d'un arbre.

⁶ Définition du mot niveau : on dira qu'un article est de niveau 1, s'il a au moins une sous section et si aucune sous section n'a elle-même de sous sections. On généralise cette définition pour définir ce qu'on appellera un article de niveau n. Nous dirons que l'article représenté par la figure 12, est de niveau 3.

Remarque : cette figure se lit de gauche vers la droite et de haut vers le bas en suivant le sens des flèches, nous présentons qu'une partie du découpage, nous avons omis volontairement le découpage de la section 5 (S5) et l'annexe 1 (A1).

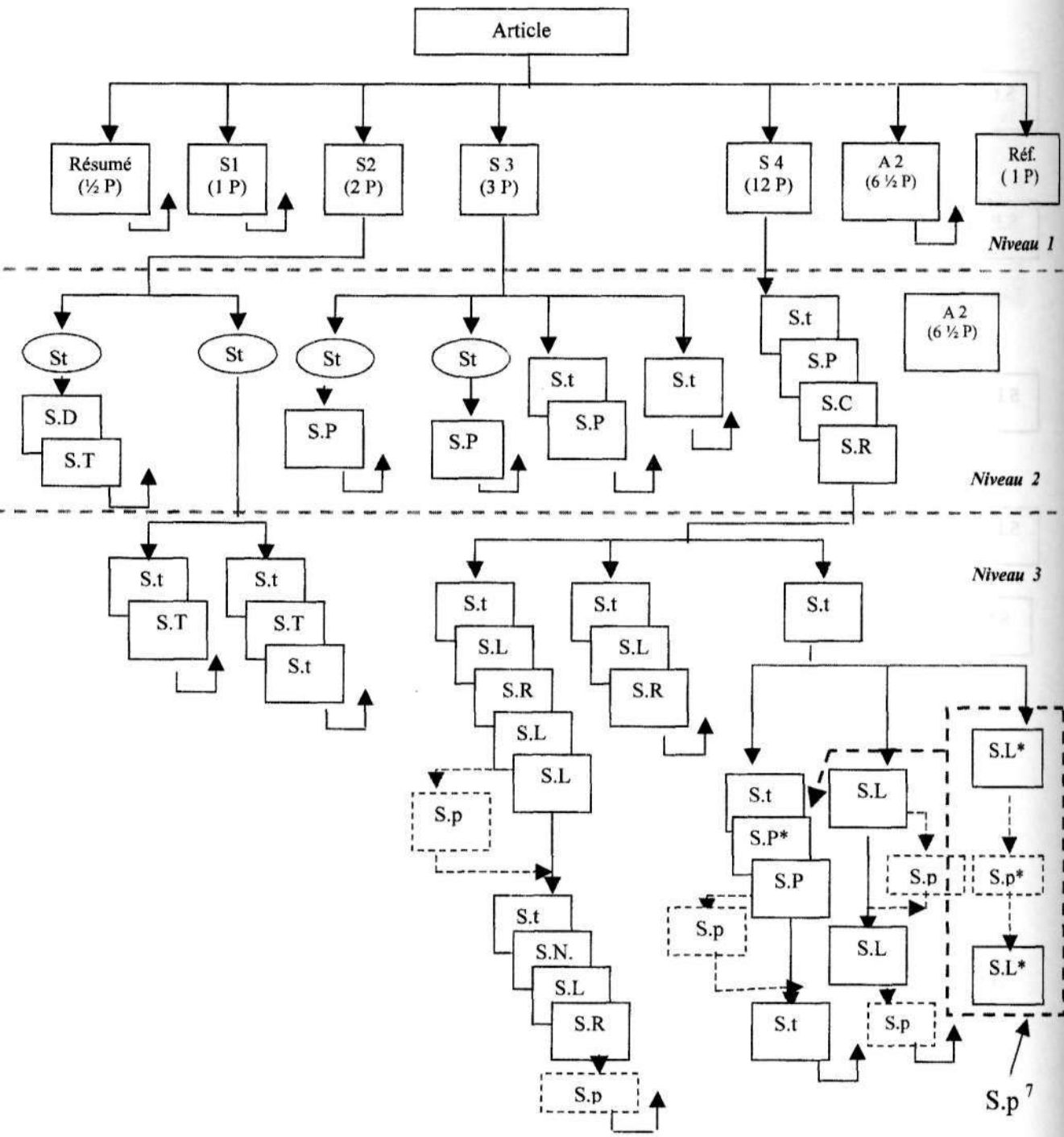


Figure 12 : Résultat du découpage selon les spécificités mathématiques

Nous avons identifié les segments suivants:

- Les segments étiquettes de niveau 1 sont comme suit : « *S.r* : *Segment résumé*, *S.I* : *Segment Introduction*, *S.T* : *Segment Théorème*, *S.L* : *Segment Lemme*, *S.P* : *Segment Proposition*, *S.D* : *Segment Définition*, *S.C* : *Segment Corollaire*, *S.E* : *Segment Exemple*, *S.A* : *Segment Application*, *S.N* : *Segment Notation*, *S.R* : *Segment Remarque*, *S.Réf.* : *Segment Référence*, *S.rem.* : *Segment Remerciement* ».
- Le segment étiquette de niveau 2 : « *S.p* : *Segment Preuve ou Segment Démonstration* ».
- Les segments texte comme nous les avons définis auparavant.

Une des difficultés majeures, c'est le regroupement des segments étiquettes mathématiques de niveau 1 possédant un segment d'étiquette de niveau 2 non contiguë. La figure 12, nous montre un exemple de ce type de problème. Pour résoudre ce dernier, nous devons passer à une deuxième phase qui consiste à restructurer notre corpus et cela en regroupant les différents segments qui ont une relation entre eux implicite ou explicite. Nous devons alors regrouper les étiquettes de niveau 1 et de niveau 2, quand elles existent dans des unités autonomes, comme le cas dans cet exemple.

3. Conclusion

La méthode que nous venons de présenter s'appuie sur les spécificités du langage du mathématicien. La mise en œuvre d'un découpage d'un corpus avec les segments étiquettes spécifiées par cette méthode permettra de proposer au mathématicien des parcours différents suivant ses besoins ou préoccupations. Il pourra par exemple consulter uniquement les théorèmes d'un article ou s'intéresser à la démonstration d'un résultat particulier.

D'autre part la navigation entre segments ou entre les unités autonomes d'un article pourra se faire également à l'aide des références bibliographiques qui sont dans le texte de l'article.

5. Bibliographie

[1] **BERTRAND-GASTALDY, Suzanne.** École de bibliothéconomie et des sciences de l'information Université de Montréal Case Postale 6128, Station A Montréal, Québec, CANADA H3C 3J7 URL, 1990.

http://www.comm.uqam.ca/~LTI/CV/publi_pdf/ICO93_hermes.pdf

[2] **RENZETTI, Françoise & TÊTU, Jean-François .** (Publié dans *La communication de l'IST dans l'enseignement supérieur et la recherche, l'effet RENATER/INTERNET*, ADBS Editions, 1995 - copyrights F. Renzetti et J.F. Têtu.).

[3] **ROLLET, Laurent.** « L'identité professionnelle d'un mathématicien » : Henri Poincaré Congrès SFHST (Lille 2001).

[http://www.univ-](http://www.univ-nancy2.fr/ACERHP/perso/rollet/textes/LIdentiteProfessionnelle.PDF)

[nancy2.fr/ACERHP/perso/rollet/textes/LIdentiteProfessionnelle.PDF](http://www.univ-nancy2.fr/ACERHP/perso/rollet/textes/LIdentiteProfessionnelle.PDF)

[4] **BERARD Pierre,** « Documentation mathématique : enjeux pour l'ère numérique » 68th IFLA Council and General Conference August 18-24, 2002. Code Number: 095-112-F, Division Number: II, Professional Group: Science and Technology Libraries
University Joseph Fourier, Grenoble, France.

[5] **TOUMIT, Jean-Yves.** « Rétroconversion de documents mathématiques », Thèse de doctorat, sous la direction de H. EMPTOZ. INSA 2000 Lyon.

[6] **BESSONNAT, D.** Le découpage en paragraphes et ses fonctions, *Pratiques* N°57, mars 1998, pp. 81-105.

[7] **OUERFELLI, T.** « La segmentation des documents techniques composites dans une perspective d'indexation : vers la définition d'un modèle dans une optique d'automatisation », sous la direction de G. Lallich. Thèse de doctorat. Juillet 2001.

[8] **DESGRAUPES, Bernard**. LATEX, Apprentissage, guide et référence. Vuibert, Paris, 2000. ISBN 2-7117-8658-7.

Les exemples que nous avons choisis ont été extraits des articles suivants :

Les figures 7, 12 ont été extraites de l'article de M. **BELLASSOUED Mourad**. « Unicité et contrôle pour le système de lame » ESAIM: Control, Optimisation and Calculus of Variations September 2001, Vol. 6, 561-592 URL: EDP Sciences, SMAI 2001

<http://www.emath.fr/cocv/>

Les figures 1, 2, 4, 5, 8, 10, 11 ont été extraites de la note de M. **TALBI**. « Inégalité de Haagerup, groupoïdes et immeubles euclidiens », C. R. Acad. Sci. Paris, Ser. I 335 (2002) 233-236, Analyse harmonique/*Harmonic Analysis*, Institut Girard Desargues, bâtiment 101, 43, bd du 11 Novembre 1918, 69622 Villeurbanne, France.

La figure 6 a été extraite de l'article de M. **CABANAL-DUVILLARD, Thierry**. « Fluctuations de la loi empirique de grandes matrices aléatoires », *Ann. I. H. Poincaré – PR* 37, 3 (2001) 373-402, 2001 Éditions scientifiques et médicales Elsevier SAS. All rights reserved. S0246-0203(00)01071-2/FLA, UFR Math-Info. Université René Descartes, rue des Saints-Pères, 75270 Paris cedex 6, France.

La figure 9 a été extraite de l'article de M. **CHAMBERT-LOIR Antoine**. « Points de petite hauteur sur les variétés semi-abéliennes », *Ann. Scient. Éc. Norm. Sup.*, 4e série, t. 33, 2000, p. 789 à 821.

Annales scientifiques de l'école normale supérieure. – 0012-9593/00/06/0 2000 Éditions scientifiques et médicales Elsevier SAS. All rights reserved.

La figure 12, est le résultat de notre découpage. Cet article est de Mr. **BELLASSOUED, Mourad**. Intitulé « Unicité et contrôle pour le système de lame », Septembre 2001, Vol. 6, 561-592, ESAIM: Control, Optimisation and Calculus of Variations, © EDP Sciences, SMAI 2001

URL: <http://www.emath.fr/cocv/>.