



**HAL**  
open science

## Les bases de données bibliographiques vecteur de connaissances de la production scientifique : étude des sciences médicales en Afrique

Alain Collignon, Pascal Cuxac

### ► To cite this version:

Alain Collignon, Pascal Cuxac. Les bases de données bibliographiques vecteur de connaissances de la production scientifique : étude des sciences médicales en Afrique. 1st. International Symposium ISKO-Maghreb, May 2011, Hammamet, Tunisie. hal-00952826

**HAL Id: hal-00952826**

**<https://hal.science/hal-00952826>**

Submitted on 27 Feb 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Les bases de données bibliographiques vecteur de connaissances de la production scientifique : étude des sciences médicales en Afrique

Alain Collignon ; Pascal Cuxac

INIST-CNRS, 2 allée du Parc de Brabois, CS 10310, 54519 Vandoeuvre les Nancy Cedex, FRANCE

prénom.nom@inist.fr

**Résumé**—Les bases de données bibliographiques sont utilisées par les chercheurs pour effectuer des recherches documentaires, faire un état de l'art, etc. Mais d'autres applications sont possibles afin de faire émerger de la connaissance sur l'état de la production scientifique d'un pays ou d'un continent dans un domaine donné.

L'accès aux résultats de la recherche et à l'information scientifique et technique est en train de changer de paradigmes et ce grâce au « libre accès » et plus particulièrement au phénomène d'auto-archivage des travaux de recherche dans des archives ouvertes.

Face à ce constat, la présente communication situe l'INIST-CNRS et la base PASCAL avec sa composante en sciences médicales comme un outil de valorisation de la production scientifiques, issue des pays Africains et d'offrir des services gratuits concernant l'information sanitaire.

Nous montrerons également comment un outil d'interrogation et d'analyse de l'information comme la plateforme Stanalyst permet d'analyser finement les publications scientifiques et d'être un outil pour un large public allant du documentaliste au décideur en passant par le veilleur et le scientifique (ou le praticien).

**Index Terms**—Databases, Scientific publishing, Bibliometrics, Africa

## I. INTRODUCTION

L'INIST est une unité de service du Centre National de la Recherche Scientifique (CNRS) qui collecte, analyse et diffuse les résultats de la recherche mondiale en sciences, technologie, médecine ainsi qu'en sciences humaines, sociales et économiques.

Hérité des anciens centres de documentation du CNRS, cet institut, créé en 1988, est localisé dans l'est de la France sur le technopôle de Nancy.

Dès le début de sa création, l'INIST-CNRS a été imaginé comme une usine de l'information : l'activité documentaire s'organise selon un processus de production qui permet de traiter d'importants volumes de données et de répondre aux exigences qualité des utilisateurs (Dusoulier, 1993). Ce processus permet, à partir d'un document primaire, d'élaborer

de nombreux produits et services. L'INIST-CNRS produit d'une part, des bases de données bibliographiques PASCAL et FRANCIS signalant respectivement 18 millions de références en sciences, technologie et médecine depuis 1973 et près de 3 millions de références en sciences humaines et sociales depuis 1972, et d'autre part, fournit la copie à la demande des documents recherchés.

Dès 2000, avec l'arrivée de services de diffusion de versions électroniques de revues, l'INIST-CNRS s'oriente vers l'acquisition et la mise à disposition de ces nouvelles formes de publication à travers des portails documentaires généralistes ou thématiques. Un service d'édition électronique est également créé à cette époque, dans le cadre de la diversification des activités d'offres de service personnalisés et intégrés pour les utilisateurs de l'Institut.

Nous verrons dans cet article quel rôle ou influence peut avoir la base de données bibliographique PASCAL dans la diffusion et la valorisation de la production scientifique africaine en sciences médicales, plus particulièrement dans le domaine de la médecine tropicale. Nous commencerons par présenter les activités autour de la production de la base PASCAL, puis nous poursuivrons avec une étude infométrique portant sur la production 2009 relative aux publications africaines en sciences médicales. Nous verrons ensuite comment les publications dans le domaine de la médecine tropicale sont valorisées dans PASCAL. Tout au long de cette étude infométrique, nous étudierons différents paramètres permettant de mettre en lumière le rôle de la base de données dans le domaine des sciences médicales vis-à-vis de l'Afrique.

## II. LA BASE DE DONNÉES PASCAL

### A. Généralités

Pour réaliser ses missions l'INIST-CNRS s'appuie sur un fonds documentaire unique qui couvre l'ensemble de la littérature cœur en sciences, technologie, médecine et sciences humaines et sociales (Guichard, 1999).

Ce fonds est essentiellement constitué par 19 700 titres de périodiques dont 4 500 abonnements en cours, 125 000 thèses

en sciences et technique, 115 000 conférences, 75 000 rapports et 13 000 monographies. Le stockage est réalisé sous forme papier représentant 25 kilomètres de rayonnage en bibliothèque et, pour des raisons de conservation, au format électronique représentant plus de 7 millions d'articles numérisés en format image (TIFF) depuis 1990.

La production de la base de données PASCAL est placée sous la responsabilité du département d'ingénierie de l'information scientifique. Ce département, est divisé en quatre services scientifiques et un service dédié à la gestion et au suivi de production. Il est composé de 56 ingénieurs documentalistes, spécialistes dans leur domaine d'indexation. Quelques coopérants, comme par exemple le réseau documentaire d'information en sciences publiques - BDSP<sup>1</sup>, viennent renforcer cette activité. Plus de 3 000 titres de périodiques, majoritairement en langue anglaise, sont analysés annuellement, correspondant à un accroissement de 480 000 références bibliographiques. La base PASCAL, interrogeable en français, anglais et espagnol est distribuée via différents supports comme des serveurs (Questel, Dialog, Datastar, ...) ou des fournisseurs d'information tels que Ovid ou EBSCO. Depuis 1973, date de la création de PASCAL, l'INIST-CNRS possède plus de 18 millions de références réparties dans différents domaines scientifique (figure 1).

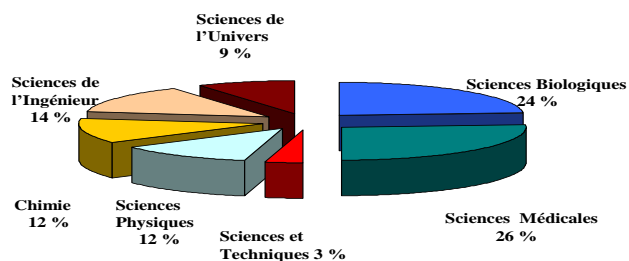


FIGURE 1 : RÉPARTITION DES DOMAINES SCIENTIFIQUES DANS PASCAL.

### B. L'indexation

Le travail d'indexation réalisé par les ingénieurs documentalistes comprend différentes tâches permettant de représenter le contenu des articles analysés par des :

- descripteurs contrôlés choisis dans les vocabulaires spécifiques;
- mots-clés non contrôlés (candidats descripteurs) ;
- codes de classement choisis dans les plans de classement élaborés par l'INIST-CNRS.

L'indexation implique deux tâches intimement liées :

- l'indexation proprement dite ;
- la constitution et la gestion des ressources nécessaires à l'indexation :
  - langages documentaires :
    - vocabulaires de référence ;
    - plans de classement ;
  - référentiels de règles de pré-indexation.

<sup>1</sup> BDSP : Banque de Données en Santé Publique ; <http://www.bdsp.ehesp.fr>

L'indexation peut-être **soit manuelle**, dans ce cas les descripteurs et les codes de classement sont saisis par le rédacteur au niveau d'une zone de saisie ou directement dans une grille d'indexation (Ménillet, 1992), **soit assistée**. Pour cette dernière, deux moteurs d'indexation sont utilisés, l'un pour réaliser la pré-indexation (approche symbolique ou linguistique faisant appel à un programme qui utilise des règles de reconnaissance de chaînes de caractères dans les textes analysés pour produire une liste d'éléments d'indexation), et l'autre la collocation lexicale (approche statistique utilisant un dictionnaire d'association entre mots du titre et éléments d'indexation ; ce dictionnaire est constitué automatiquement lors d'un processus d'apprentissage sur un corpus indexé).

### C. Les sciences médicales dans PASCAL

Comme nous avons pu le lire précédemment, la base PASCAL regroupe différents domaines scientifiques, dont les sciences médicales qui y occupent une place importante. En 2009, près de 1 400 titres de périodiques en médecine, subdivisés en 6 grandes spécialités médicales (figure 2) ont été acquis par l'Institut.

	Nb. Titres (en 2009)	Nb. Articles / an (stats sur 3 ans)	Volumétrie (1984-2009)
<b>Médecine générale</b>	<b>130</b>	<b>15 771</b>	<b>323 783</b>
<b>Imagerie Techniques d'exploration</b>	<b>39</b>	<b>7 379</b>	<b>154 333</b>
<b>Santé publique</b>	<b>61</b>	<b>4 699</b>	<b>76 613</b>
<b>Spécialités médicales</b>	<b>628</b>	<b>112 994</b>	<b>2 078 367</b>
<b>Pharmacologie et toxicologie</b>	<b>133</b>	<b>24 053</b>	<b>416 658</b>
<b>Psychologie, psychiatrie</b>	<b>309</b>	<b>23 430</b>	<b>359 406</b>

FIGURE 2 : RÉPARTITION DES SCIENCES MÉDICALES DANS PASCAL EN 2009

Comme pour l'ensemble du fonds documentaire, la langue de publication majoritaire est l'anglais à l'exception des documents publiés sur le continent africain pour lesquels le français y est utilisé à près de 75 %. Concernant les publications issues des pays d'Amérique du Sud, les langues portugaise et espagnole y sont majoritaires à 38% chacune (figure 3).

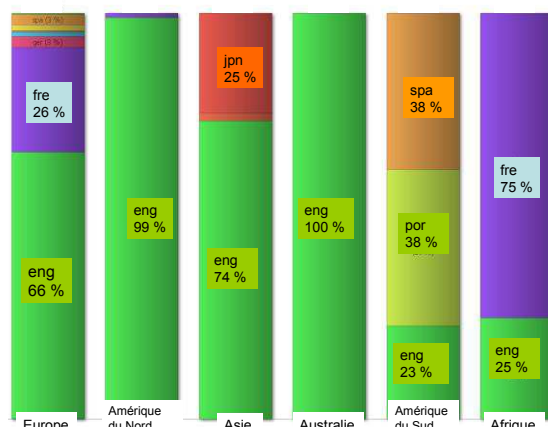


FIGURE 3 : VENTILATION DES TITRES EN SCIENCES MÉDICALES PAR CONTINENT ET PAR LANGUE DE PUBLICATION

Après avoir présenté la base de données PASCAL et sa composante en sciences médicales, nous allons maintenant exposer l'analyse scientométrique de ce domaine scientifique publié sur le continent africain à l'aide de l'outil Stanalyst, et en développant plus précisément le sous-domaine « Médecine Tropicale ».

### III. ETUDE SCIENTOMÉTRIQUE DES SCIENCES MÉDICALES PUBLIÉES SUR LE CONTINENT AFRICAÏN

#### A. Le corpus étudié

Notre étude s'est limitée à analyser l'ensemble des notices bibliographiques signalées en 2009 dans la base PASCAL en sciences médicales. Le corpus pris en compte n'est qu'une sous-partie de cette base ayant fait l'objet d'une étude bibliométrique sur l'état de la Science produite en Afrique (Arvanitis, et al. 2000).

Le corpus étudié contient plus de 150 000 notices (pour 2 050 périodiques) publiées dans le monde dont environ 65 500 (pour 1500 périodiques) par le continent européen et près de 4 000 notices (pour 646 périodiques) par le continent africain. Cette étude permet d'avoir une vue instantanée sur la production en sciences médicales produite sur le continent africain. Une étude comparative sur plusieurs années est en cours d'analyse.

#### B. La méthode d'analyse scientométrique utilisée

Les corpus, issus de la base de données bibliographique PASCAL, sont créés et analysés en utilisant la station d'analyse Stanalyst (Grivel et al. 1995) ; <http://stanalyst.inist.fr>). Cette station regroupe quatre modules (figure 4) permettant de créer la requête d'interrogation des bases de données, de faire une analyse bibliométrique des résultats, de filtrer les termes d'indexation et enfin de faire une classification non supervisée qui sera ensuite représentée sous forme de carte.

Le corpus issu de l'interrogation de la base de données PASCAL est constitué de notices bibliographiques structurées en champs (figure 5).

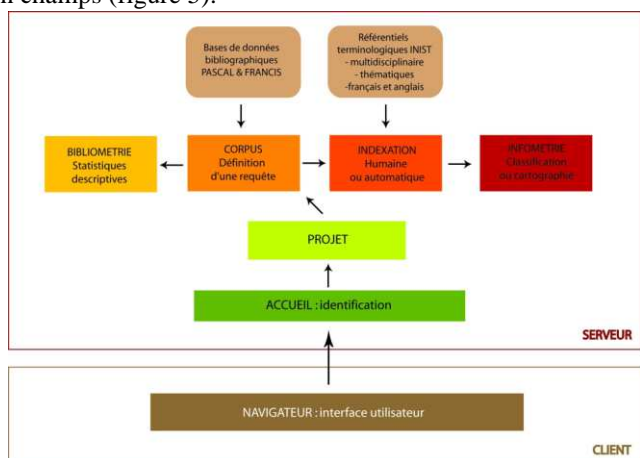


FIGURE 4 : ARCHITECTURE DE LA PLATEFORME STANALYST

Le module « bibliométrique » de Stanalyst permet de faire des statistiques descriptives sur plusieurs de ces champs

(auteurs, pays d'affiliation, pays d'édition, langues, descripteurs, etc.).

**ET** : Epilepsie par encornement bovin (zébu): à propos d'un cas au Burkina Faso  
**ET** : (Epilepsy by bovine (zebu) goring: a case report in Burkina Faso)  
**AU** : NAPON (C.); DRAVE (A.); KABORE (J.)  
**AF** : Service de neurologie du CHU Yalgado-Ouedraogo, BP 7022 Ouagadougou 03/Burkina Faso (1 aut., 2 aut., 3 aut.)  
**DT** : Publication en série; Niveau analytique  
**SO** : Bulletin de la Société de pathologie exotique; ISSN 0037-9085; Coden BSPEAM; France; Da. 2009; Vol. 102; No. 4; Pp. 217-218; Abs. anglais; Bibl. 4 ref  
**LA** : Français  
**EA** : The post-traumatic epilepsy is responsible for 20% of the symptomatic epilepsies. Accidents on public highway constitute more than 70% of the causes. We report a singular case of fronto-polar post-traumatic epilepsy by zebu goring which appeared two years after the traumatism. The neurological ...  
**CC** : 002B01; 002B17A03; 002B16B  
**FD** : Epilepsie; Traumatisme crânien; Infection; Bovin; Etude cas; Burkina; Hôpital; Médecine tropicale; Homme  
**FG** : Artiodactyla; Ungulata; Mammalia; Vertebrata; Afrique; Pathologie de l'encéphale; Pathologie du système nerveux central; Pathologie du système nerveux  
**ED** : Epilepsy; Head trauma; Infection; Bovine; Case study; Burkina Faso; Hospital; Tropical medicine; Human  
**EG** : Artiodactyla; Ungulata; Mammalia; Vertebrata; Africa; Cerebral disorder; Central nervous system disease; Nervous system diseases  
**SD** : Epilepsia; Traumatismo craneoencefálico; Infección; Bovino; Estudio caso; Burkina Faso; Hospital; Medicina tropical; Hombre

FIGURE 5 : EXEMPLE D'UNE NOTICE BIBLIOGRAPHIQUE PASCAL

Le module « indexation » permet soit de faire une indexation automatique du corpus si on possède une ressource terminologique, soit d'utiliser l'indexation native des notices. Dans tous les cas le résultat, d'indexation doit être révisé afin d'enlever les termes trop génériques ou non porteurs de sens pour l'analyse considérée, et cela dans le but d'effectuer une classification.

Cette classification est obtenue en utilisant le module « infométrie ». Il permet de réaliser deux types de classifications non supervisées (clustering) :

Neurodoc : est une méthode inspirée des K-means (MacQueen 1967). Cet algorithme, que nous avons utilisé pour analyser les différents corpus et produire des cartes thématiques, va être détaillé ci-dessous.

Sdoc : est une méthode de classification basée sur les mots associés (Callon et al., 1983)

L'algorithme des K-means consiste à créer K groupes de documents de façon à ce que chaque groupe soit le plus dense possible et que ces groupes soient les plus distants possibles. L'utilisateur choisit le nombre de classes maximum désiré (K) et va initialiser K classes par un point représentant le centre de chaque classe. Les documents sont ensuite affectés à la classe dont ils sont le plus proche et le centre de la classe est recalculé jusqu'à stabilisation.

La méthode des K-means axiales (Lelu, 1993) diffère en représentant les classes, non plus par leur centre, mais par un demi-axe. Les documents se projettent alors plus ou moins loin sur ces axes. Quand la projection se fait au dessus d'un certain seuil fixé par l'utilisateur le document est affecté à la classe. On voit ainsi qu'un document peut appartenir à plusieurs classes (classes recouvrantes). De la même façon on peut projeter les mots-clés sur ces axes. Les documents et mots-clés sont ordonnés selon un degré de ressemblance au type idéal de la classe et sont affectés d'un poids plus ou moins important suivant qu'ils se projettent plus ou moins haut sur ces axes.

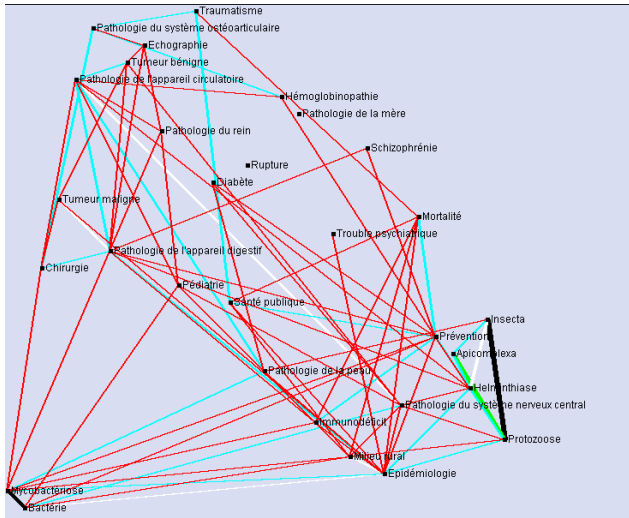


FIGURE 8 : UN EXEMPLE DE RÉSULTAT DE CLASSIFICATION

A l'issue de cette étape de classification, les documents sont représentés dans l'espace de description, c'est-à-dire dans l'espace des mots-clés (il n'est pas rare d'avoir des corpus représentés dans des espaces à 5 000 dimensions).

Pour obtenir un résultat lisible par l'œil humain nous réalisons une analyse en composantes principales (ACP) afin de représenter le résultat sur un plan. Nous faisons également figurer les valeurs remarquables des cosinus entre classes, ce qui permet d'esquisser un réseau de classes (figure 8).

### C. Résultats

#### 1) La publication : qui et comment ?

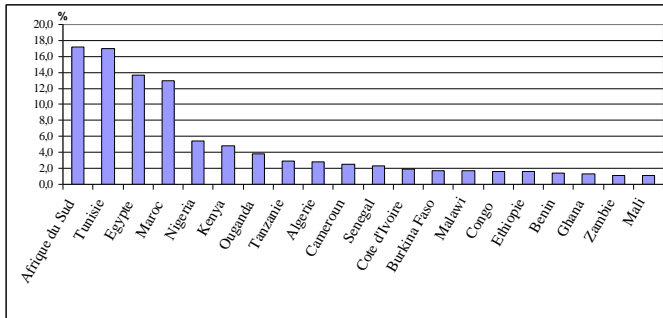


FIGURE 9 : VENTILATION DE LA PRODUCTION EN SCIENCES MÉDICALES EN FONCTION DES PAYS D'AFFILIATION

Cette figure représente les vingt premiers pays africains d'affiliation. L'Afrique du Sud et la Tunisie en sont les deux pays principaux (environ 17 % de la production), suivis de près par l'Égypte et le Maroc (environ 13%). Puis s'en suit une longue traîne allant de 4,5 % à 1 %, composée majoritairement par les pays d'Afrique subsahariennes et l'Algérie.

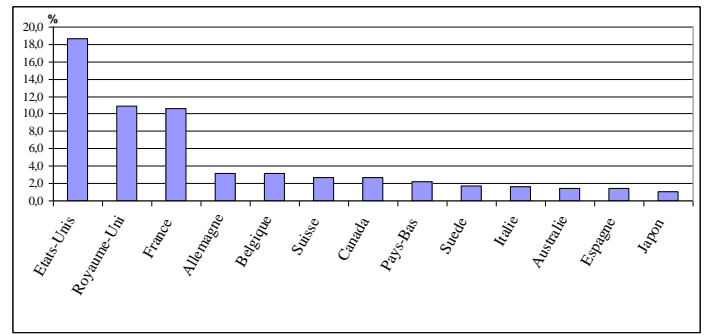


FIGURE 10 : RÉPARTITION DES CO-PUBLIANTS EN SCIENCES MÉDICALES HORS AFRIQUE

Les articles scientifiques publiés par les laboratoires appartenant à des pays africains sont rédigés en collaboration avec des collègues étrangers pour plus de 60 %. Les chercheurs issus des Etats-Unis accompagnent majoritairement (près de 20 %) les chercheurs africains. La France et le Royaume – Uni suivent à près de 10 % chacun.

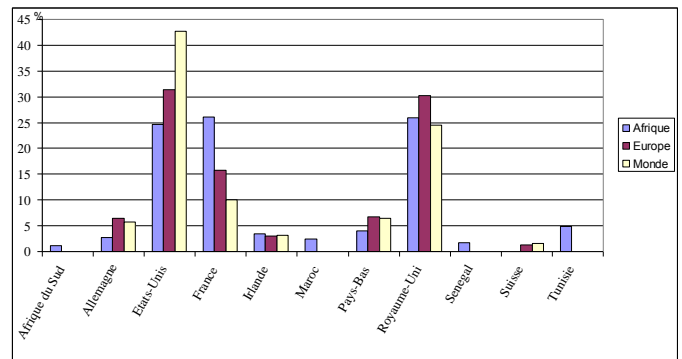


FIGURE 11 : REPRÉSENTATION DE LA LIAISON PAYS D'ÉDITION AVEC PAYS D'AFFILIATION EN SCIENCES MÉDICALES

La figure 11, illustre la répartition des pays d'édition les plus représentatifs combinés avec les pays de publication en sciences médicales. Nous constatons, sans grande surprise, que les périodiques publiés dans les journaux américains et anglais sont majoritairement composés d'articles rédigés par des chercheurs soit européens ou du reste du monde. Tandis que, dans les journaux publiés en France, nous constatons une proportion non négligeable d'articles rédigés par les chercheurs africains. La présence de publications dans ces journaux est sans doute due aux divers problèmes liés à la diffusion. En revanche pour les journaux édités par les pays africains (Afrique du Sud, Maroc, Sénégal et Tunisie pour ne citer qu'eux), nous retrouvons exclusivement des articles publiés par des chercheurs africains.

Il est aussi démontré que les chercheurs s'intéressant au domaine étudié, réservent leurs travaux pour Elsevier (pour 18 % des cas), suivis des éditeurs Springer et Blackwell (7 % chacun). Le reste de la production est diffusé via un bouquet d'éditeurs pour une faible proportion chacun (1 %) et de sociétés savantes.

Pour conclure cette partie, nous pouvons dire que notre étude bibliométrique réalisée sur un corpus de plus de 4 000 notices bibliographiques dans le domaine des sciences médicales publiées par au moins un chercheur africain, nous indique que l’Afrique du Sud et la Tunisie jouent un rôle essentiel avec à leur cotés le Maroc et l’Egypte et que les pays d’Afrique subsaharienne ainsi que l’Algérie participent activement à la production scientifique des sciences médicales. Même si la coproduction Afrique – Monde est significative (dépendante des programmes d’aide à la recherche ou de coopération financés par le Nord), la coproduction intra pays africains n’est pas faible signifiant une volonté accrue d’augmenter la visibilité de leur propre travaux de recherche. Les chercheurs africains se tournent volontiers vers les Etats-Unis et le Royaume-Uni comme pays d’édition, ainsi que vers la France. D’ailleurs, ce sont dans les périodiques publiés en France que nous retrouvons une part importante de documents produits par l’Afrique. Par conséquent, ces chercheurs ont le souci de conserver la langue française comme vecteur de la science (32% des notices du corpus sont en langue française). Néanmoins, la publication régionale (qui doit correspondre à des thématiques d’intérêt local) n’est pas négligeable ce qui est encourageant pour l’avenir.

2) *Résultat de classification du corpus « sciences médicales » et représentation cartographique*

Deux interrogations de la base de données PASCAL ont permis d’obtenir les corpus suivants :

- un sur la recherche médicale en Europe en 2009 (un auteur au moins issu d’un pays européen),
- un sur la recherche médicale en Afrique en 2009 (un auteur au moins issu d’un pays africain),

Les résultats présentés ci-dessous (figures 12 et 13) ont été obtenus par la méthode des K-means axiales en initialisant avec 50 classes.

a) *La recherche médicale en Europe :*

Une analyse rapide de cette carte (figure 12) permet de dégager quelques grands pôles de recherche. En partant du haut de la carte et dans le sens horaire on trouve : la psychiatrie, les études sur le système nerveux, la cardiologie, l’anesthésie, l’obésité, le cancer, le sida, la rhumatologie/orthopédie. Bien entendu une analyse plus détaillée de cette carte permet d’obtenir plus d’informations sur le domaine étudié mais ce n’est pas le but de cette publication.

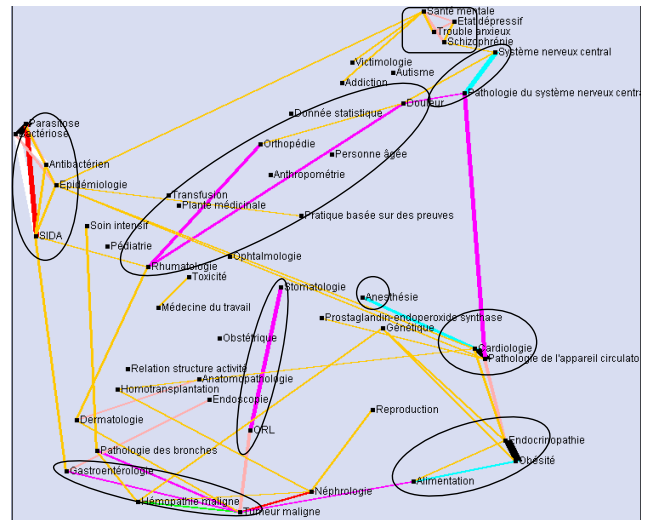


FIGURE 12 : RÉSULTAT DE LA CLASSIFICATION SUR LE CORPUS « MÉDECINE – EUROPE »

La classe la plus importante en nombre de documents (Tumeur maligne) rassemble près de 12 % des documents, suivie de « Pathologie de l’appareil circulatoire » et « Pathologie du système nerveux central » avec chacune 9 % des documents.

b) *La recherche médicale en Afrique :*

En première approche on voit se dégager les pôles de recherches suivants (figure 13) : Chirurgie ; Tabagisme ; Pharmacognosie ; Sida ; Médecine tropicale.

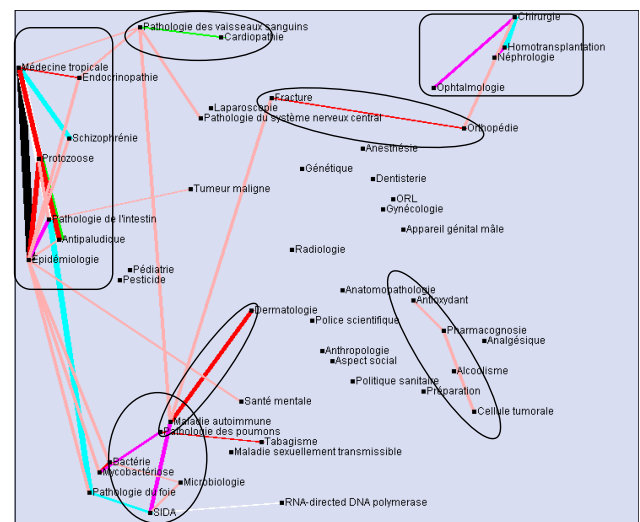


FIGURE 13 : RÉSULTAT DE LA CLASSIFICATION SUR LE CORPUS « MÉDECINE – AFRIQUE »

La classe la plus importante est sans conteste « Médecine tropicale » avec près de 22 % des documents du corpus. Viennent ensuite les classes « Sida » (13 %), « Epidémiologie » (12%) et « Protozoose » (8 %).

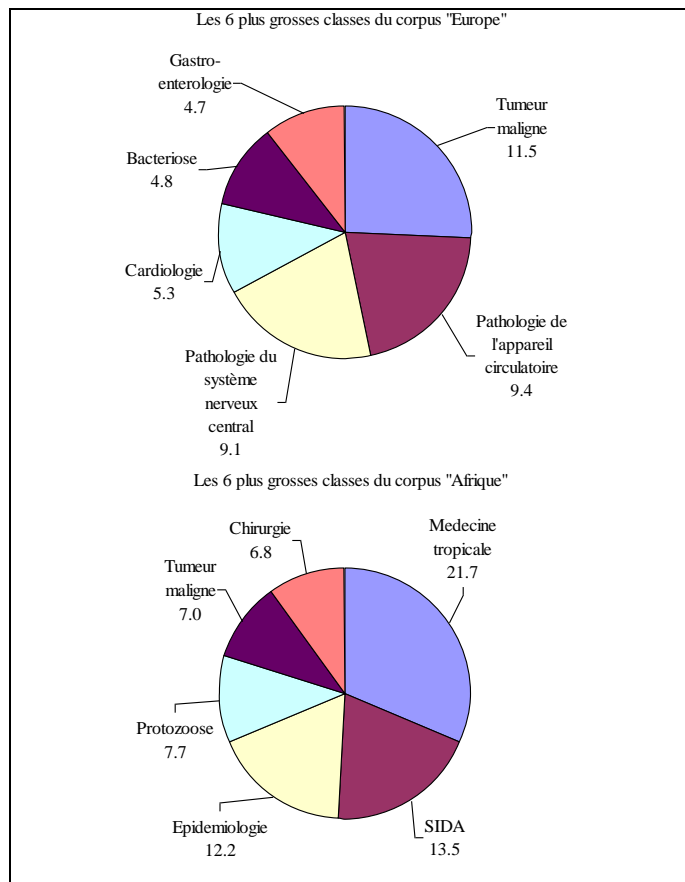


FIGURE 14 : RÉPARTITION DES CLASSES LES PLUS VOLUMINEUSES POUR LES CORPUS MÉDECINE « EUROPE » ET « AFRIQUE »

Il est intéressant de comparer ces deux cartes : les classes les plus importantes sont radicalement différentes (figure 14). Pour rendre plus visible, dans cette figure ne sont illustrées que les six classes les plus importantes pour chaque classification : Cancer, Cardiologie, Système nerveux pour le corpus 'Europe' alors que dans le corpus 'Afrique' est constitué de Médecine tropicale, Sida, Epidémiologie et Protozoose.

### 3) Etude de la discipline « Médecine tropicale »

La carte des thèmes en sciences médicales produite par les chercheurs africains met en exergue la médecine tropicale (voir figure 13). Nous avons décidé de réaliser un focus sur cette discipline. Notre analyse porte sur un corpus de plus de 900 notices bibliographiques correspondant à près de 20 000 auteurs.

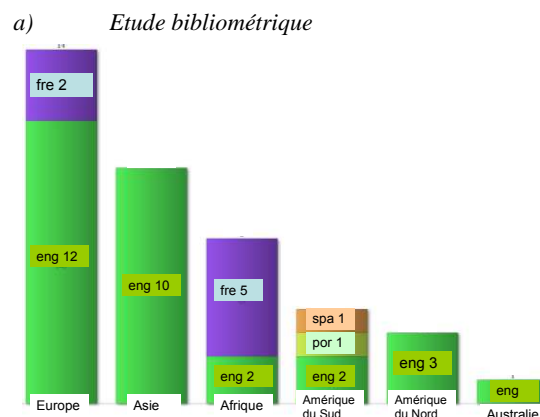


FIGURE 15 : RÉPARTITION PAR ZONE GÉOGRAPHIQUE ET PAR LANGUE DE PUBLICATION

En 2009, 38 titres en médecine tropicale sont présents dans le fonds documentaire de l'INIST. Les titres proviennent majoritairement d'Europe et sont publiés en anglais, néanmoins 15 % des revues sont originaires d'Afrique. Pour ce continent, les articles sont publiés majoritairement en français (figure 15).

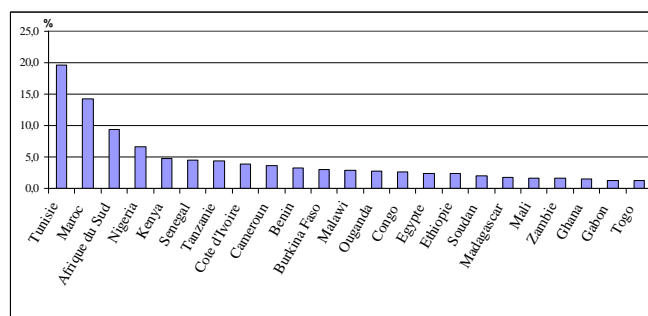


FIGURE 16 : RÉPARTITION PAR PAYS D'AFFILIATION DES PUBLICATIONS EN MÉDECINE TROPICALE

Environ 50 % de la production est produite par quatre pays qui sont par ordre de production : la Tunisie, le Maroc, l'Afrique du Sud, le Nigéria (figure 16). Nous constatons, que cet ordre est différent par rapport à l'ordre des pays producteurs en sciences médicales. Effectivement, dans cette thématique, l'Afrique du Sud et l'Égypte étaient deux pays bien représentés.

En médecine tropicale, la publication intra-Afrique est plus élevée que la publication en coopération avec le monde. Plus de 50% des publications en médecine tropicale font l'objet d'une coopération entre pays africains. Ce taux était seulement de 40% pour les sciences médicales.

### b) Résultat de classification du corpus « Médecine tropicale – Afrique » et représentation cartographique

Nous avons ici une carte avec beaucoup moins de liens forts (figure 17) : cela traduirait-il une recherche plus « éclatée » ? On note cependant quelques pôles importants. Avec les classes « Insecta » « Apicomplexa » « Heminthiase » « Protozoose » nous avons entre autres les notices portant sur le paludisme et autres parasitoses ; avec les classes « Mycobactériose »

« Bactérie » nous avons des notices portant sur la tuberculose et autres infections (tétanos).

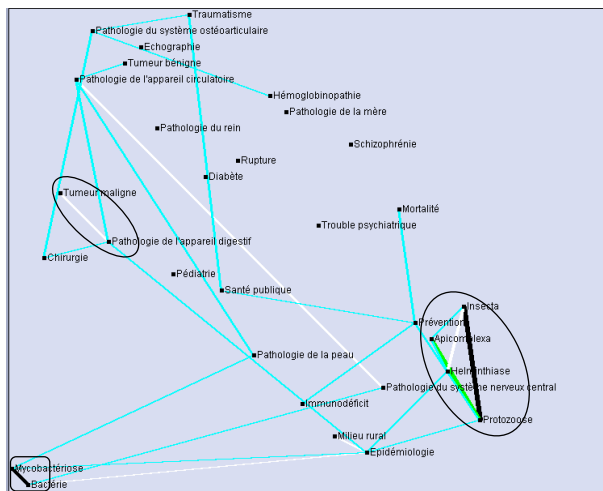


FIGURE 17 : RÉSULTAT DE LA CLASSIFICATION SUR LE CORPUS « MÉDECINE TROPICALE – AFRIQUE »

Les classes les plus importantes (en termes de volume de publications) sont « Protozoose », « Helminthiase », « Immunodéficit » (figure 18).

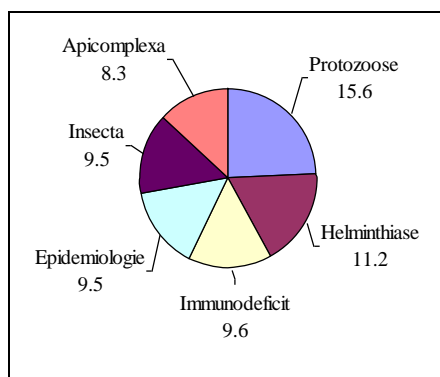


FIGURE 18 : RÉPARTITION DES CLASSES LES PLUS VOLUMINEUSES POUR LE CORPUS MÉDECINE TROPICALE « AFRIQUE »

#### IV. CONCLUSION

L'Afrique est constituée d'une cinquantaine de pays, mais les sept plus importants représentent près de 80% de la production scientifique du continent, et deux pays (l'Afrique du sud et l'Egypte) produisent la moitié des publications africaines. Les pays les plus visibles scientifiquement (Afrique du Sud, Egypte, Maroc, Tunisie, Kenya, Nigeria, Algérie) sont éloignés les uns des autres ou séparés par des frontières qui ne facilitent pas les relations entre eux. Néanmoins, avec ce potentiel, l'Afrique produirait un peu plus de 0,3% de la production scientifique mondiale.

Les résultats de notre étude concernant les sciences médicales produites par au moins un chercheur africain, en particulier pour la spécialité de la médecine tropicale, et analysées dans la base de données PASCAL sont le reflet de ce constat.

Nous avons vu que la plateforme Stanalyst est un outil à la

fois d'interrogation des bases de données et d'analyse des corpus obtenus (statistiques descriptives et classification). Stanalyst permet l'interrogation des bases de données de l'INIST mais il pourrait également être adapté pour travailler sur d'autres bases de données bibliographiques. Dans le cadre d'une mutualisation des données (base de données d'informations sanitaires africaine ?), cet outil permettrait à tout un chacun une interrogation simple mais surtout une analyse fine des résultats pouvant aider à la fois le documentaliste, le praticien, le chercheur, ou encore le décideur.

Les bases de données bibliographiques sont un vecteur de connaissance permettant d'appréhender l'état de la recherche (domaines étudiés, production scientifique, coopérations internationales...). En tant que telles, elles sont une vitrine de la recherche nationale, d'où la nécessité selon nous que des pays valorisent leurs résultats scientifiques via des bases de données qui pourraient être alimentées à l'échelle d'un continent (l'Afrique dans le cas traité ici).

Cette étude dont la problématique est d'étudier la place de la production scientifique des pays du Sud au sein de la communauté mondiale, pourrait être poursuivie par un questionnement sur l'accès à l'information scientifique et technique par les chercheurs du Sud. En effet, nous savons tous que l'accès aux revues prestigieuses est hors de portée pour la majorité des universités et centres de recherches car ils sont dans l'impossibilité de financer ces abonnements trop onéreux. Différents facteurs comme par exemple l'initiative HINARI<sup>2</sup>, le projet SIST<sup>3</sup>, ou bien encore les archives ouvertes et leurs accès par différents moissonneurs pourront être analysés. Ces éléments d'études seront étudiés comme une chance pour la recherche africaine.

#### V. BIBLIOGRAPHIE

- [1] Arvanitis R., Waast R., Gaillard J. C. (2000), "Science in Africa : A bibliometric panorama using PASCAL database", *Scientometrics*, No 47-3, pp. 457-473.
- [2] Dusoulier, N. (1993), « L'INIST, au cœur de la recherche et de l'Europe », *Documentaliste- Sciences de l'Information*, No 30-1, pp. 19-22.
- [3] Grivel L., François C. (1995) : Une station de travail pour classer, cartographier et analyser l'information bibliographique dans une perspective de veille scientifique et technique", *Solaris* n° 2, 1995, p. 81-112 ; <http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d02/2grivel.html>
- [4] Guichard, M. (1999), « L'INIST-CNRS: des services d'information payants et gratuits dans un cadre de service public », *Bulletin d'Informations ABF*, No.184-5, pp.100-2.
- [5] MacQueen J. B. (1967) : Some Methods for classification and Analysis of Multivariate Observations, *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1: 281-297.
- [6] Ménillet, D. (1992), « Grilles d'indexation et de préindexation. L'exemple de PASCAL », *Documentaliste-Sciences de l'Information*, No. 29-4/5, pp.183-190.
- [7] Lelu A. (1993) : Modèles neuronaux pour l'analyse de données documentaires et textuelles, Thèse de doctorat de l'université de Paris VI, 4 mars 1993, 238 pages.
- [8] Callon M., Courtial J-P., Turner W. A., Bauin S. (1983) : From Translation to Problematic Networks: An Introduction to Co-Word Analysis, in *Social Science Information*, vol. 22, pp. 191-235.

<sup>2</sup> <http://www.who.int/hinari/fr/index.html>

<sup>3</sup> <http://www.sist-sciencesdev.net/>