

1988  
DSB  
23

ECOLE NATIONALE SUPERIEURE DE BIBLIOTHECAIRES  
DIPLOME SUPERIEUR DE BIBLIOTHECAIRE  
OPTION:  
CONCEPTION ET GESTION DES SYSTEMES ET RESEAUX D'INFORMATION

**SYSTEMES DOCUMENTAIRES  
MULTILINGUES**



PROJET D.S.B.

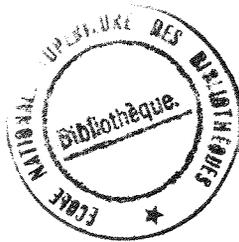
réalisé par : Maria Lina PEREIRA E SILVA  
sous la direction de : Mlle. SYLVIE LAINE

1988  
DSB  
23

1987/1988

ECOLE NATIONALE SUPERIEURE DE BIBLIOTHECAIRES  
DIPLOME SUPERIEUR DE BIBLIOTHECAIRE  
OPTION:  
CONCEPTION ET GESTION DES SYSTEMES ET RESEAUX D'INFORMATION

**SYSTEMES DOCUMENTAIRES  
MULTILINGUES**



PROJET D.S.B.

réalisé par : Maria Lina PEREIRA E SILVA

sous la direction de : Mlle. SYLVIE LAINE

1988  
DSB  
23

1987/1988

## SOMMAIRE

1. Introduction .....	1
. Justification du sujet.....	4
2. Méthodologie de recherche documentaire.....	5
2.1 Recherche manuelle.....	5
. Laboratoire d'Informatique Documentaire.....	5
. Revues de sommaires.....	5
2.2 Recherche automatisée.....	6
. Choix des Bases de Données.....	7
. Stratégie de recherche.....	8
3. Résultats et commentaires.....	10
4. Repérage des documents.....	12
5. Bibliographie.....	13



## 1. Introduction

Le multilinguisme est une réalité vécue quotidiennement par les institutions européennes et de plus en plus de problèmes de communication et d'échange regional et international d'information se présentent dans les domaines scientifique et technique, rendant nécessaire le développement d'outils de communication efficaces.

Pour mettre à disposition d'un nombre toujours croissant d'utilisateurs ces outils et de nouveaux moyens de surmonter ces difficultés , plusieurs systèmes multilingues ont été conçus.

Un système documentaire multilingue est un système dans lequel le contenu de documents rédigés dans une langue A est exploité et décrit par un indexeur, tandis qu'un utilisateur en formulant des questions dans une autre langue, par exemple la langue B, pourra obtenir des réponses en langue B, aussi bien qu'en langues A et B. Les caractéristiques d'un système multilingue sont:

- possibilité d'indexation dans plus d'une langue
- possibilité de formulation des questions et de recherche automatique dans plus d'une langue et, en conséquence,
- possibilité d'obtention des résultats dans plus d'une langue. (29)

Ce Projet, dans le cadre du Diplôme Supérieur de Bibliothécaire, a comme objectif de délimiter les lignes de recherche à suivre:

### a. Evolution de ces systèmes:

L'initiative de l'Institut Textile de France, avec la création du Système TITUS a montré la viabilité des efforts entrepris. Le Système TITUS produit des résumés traduits en quatre langues: français, anglais, allemand et espagnol.

Ensuite d'autres systèmes sont apparus à savoir:

- le DIRR - Documentation Internationale de Recherche Routière en trois langues: allemand, anglais et français;

- le SABIR-CANCERNET - du l'Institut Gustave Roussy, France, en six langues: français, anglais, allemand, tchèque, polonais et espagnol;

- le SDIM, en métallurgie, en français, anglais et allemand; et,

- le PRECIS - Preserved Context Indexing System - conçu à l'origine à la British National Bibliography comme un système d'indexation par matière, en anglais, et utilisé par d'autres organismes et dans d'autres langues européennes. (29)

### b. Développement des techniques d'élaboration de thesauri multilingues:

Les problèmes d'accès aux références bibliographiques dans les différentes langues de la Communauté Européenne, ont été pris en compte par EURONET dont le but était de créer une plus grande coopération en Europe et de rendre facilement accessible au plus grand nombre d'utilisateurs européens l'information scientifique et technique.

Ce réseau a mené plusieurs actions, en particulier l'élaboration de vocabulaires standardisés multilingues. On

ajoute ici, le rôle de l'UNESCO/ISO au niveau international, avec l'élaboration de normes pour l'établissement et le développement de thesauri multilingues; on peut citer comme exemples:

- EUDISED - thésaurus multilingue pour le traitement de l'Information en Education, en cinq langues;

- EBC - thésaurus de l'European Brewery Convention dans le domaine de la brasserie, en quatre langues;

- SDIM - thésaurus du Système de Documentation et Information Métallurgique. (29)

**c. Le rôle que joue la traduction automatique dans ce contexte:**

Dans un système de traduction automatique, la langue source est analysée, transférée dans une structure sans ambiguïté puis synthétisée dans l'autre langue. Les principaux problèmes de ce processus sont dus à l'ambiguïté et c'est la raison pour laquelle on rencontre plusieurs tentatives de mise au point de ces systèmes.

Parmi les divers systèmes de traduction automatique déjà testés, on peut citer comme opérationnels le TAUM-METEO (anglais-français), le SYSTRAN (russe-anglais), le POLYDOC (scandinave-anglais) et le TITUS (allemand, anglais, français et espagnol). (7)

. **Justification du sujet:**

Ce sujet m'a été proposé par Mlle. Sylvie Lainé, enseignante à l'ENSB (Ecole Nationale Supérieure de Bibliothécaires) et chercheur du Laboratoire d'Informatique Documentaire de l'Université Claude Bernard - Lyon I, lié au Groupe SYDO.

Le Groupe Sydo (Systèmes Documentaires) est formé par des chercheurs du Laboratoire d'Informatique Documentaire de l'Université Claude Bernard (LYON I), du Centre de Recherches en Informatique pour les Sciences Sociales (C.R.I.S.S.) de l'Université de Grenoble et du Centre de Recherches Linguistiques et Sémiologiques de l'Université Lumière (LYON II). Ce groupe a créé des programmes qui ont pour but le développement d'une stratégie d'analyse des documents par intermédiaire d'un analyseur à niveaux morphologique et syntaxique, qui vise l'élimination des ambiguïtés qui existent en des systèmes documentaires. Tout ce travail de développement d'un modèle linguistique français cherche, entre autres, à trouver l'accès le plus rapide à des bases de données multilingues. (\*)

(\*) Informatique et Sciences Humaines: pratiques et perspectives. Editions SME, 1987. 80p.

## 2. Methodologie de Recherche Documentaire

### 2.1 Recherche manuelle

#### . Laboratoire d'Informatique Documentaire:

Le Laboratoire d'Informatique Documentaire a un petit fonds documentaire spécialisé et c'est la raison pour laquelle on a commencé la recherche par là.

Une grande quantité d'ouvrages proches du sujet en question a été trouvée, mais, en fait, un ouvrage seulement, le plus important, car il s'agit d'un congrès Européen sur les Systèmes et Réseaux Documentaires, correspondait à ce qu'on voulait.

#### . Revues de sommaires:

A la Bibliothèque de l'ENSB, des revues de sommaires et des index annuels des revues spécialisées en information, ont été cherchés, et le choix des revues "Bulletin de Bibliothèques de France", "Documentaliste" et "Documentation et Bibliothèques" a été fait, pour deux raisons. D'abord parce qu'elles sont des revues de langue française; ensuite, comme elles essaient de dépouiller la grande majorité des périodiques connus, spécialisés dans ce domaine, on aura certainement un panorama bien détaillé en ce qui concerne les travaux sur le sujet.

Le Tableau 1 donne les résultats de la recherche faite dans les revues de sommaires, en utilisant les descripteurs suivants: MULTILINGUISME, BASE DE DONNEES TEXTUELLE, TEXTE INTEGRAL, TERMINOLOGIE, RESUME, TRADUCTION AUTOMATIQUE, THESAURUS.

Revue de sommaires			
Revue	Refer. sorties	Refer. pertinents	Année
BBF	14	04	85/6
Documentaliste	02	01	87
Doc. et Bibl.	02	01	85/6
TOTAL	18	06	

Tableau I

## 2.2 Recherche automatisée:

L'interrogation des bases de données était sûrement nécessaire du fait que, les références obtenues par la recherche manuelle, restaient insuffisantes.

## **. Choix des bases de données**

La sélection des bases de données ne représentait pas un problème, car le sujet, malgré une certaine interdisciplinarité, reste bien précis sur son aspect particulier du multilinguisme.

Les bases de données choisies étaient donc:

- PASCAL SCIENCES DE L'INFORMATION
- LISA-LIBRARY INFORMATION SCIENCE ABSTRACTS.

**PASCAL SCIENCES DE L'INFORMATION** - c'est une base de références bibliographiques du CDST (Centre de Documentation Scientifique et Technique) du CNRS, dans le domaine de Sciences de l'Information, qui recense des articles de périodiques, des thèses, des actes de congrès et des rapports. Son serveur est TELESYSTEMES QUESTEL.

### **LISA-LIBRARY AND INFORMATION SCIENCE ABSTRACTS**

C'est une base de données de références bibliographiques de la Library Association, dans le domaine des Sciences de l'Information, compilée non seulement par la Library Association, mais aussi par ASLIB à partir de tous types de documents: périodiques, ouvrages, communications scientifiques et rapports. Son serveur est DIALOG.

L'interrogation a été faite à l'URFIST (Unité Régionale de Formation et de Promotion pour l'Information Scientifique et Technique) de Lyon.

## . Stratégie de Recherche

La stratégie de recherche utilisée a nécessité des modifications selon le logiciel d'interrogation. On a utilisé l'opérateur logique ET pour aboutir a un résultat moins exhaustif, mais plus précis.

### Base de données PASCAL

#### 1. MULTI?LINGUE? OU MULTI?LINGUISME?

354

#### 2. 1 ET SYSTEME DOCUMENTAIRE

32

#### 3. 1 ET SYSTEME? 1AV DOCUMENTAIRE?

11

#### 4. 2 ET 3

43

Jusqu'à la deuxième question on a recherché seulement dans le champ DESCRIPTEUR. Le résultat de la troisième question est issu d'une recherche parallèle dans le champ RESUME.

### Base de données LISA

On a utilisé la commande EXPAND de DIALOG qui permet de prendre en compte toutes les extensions possibles d'un terme, dans tous les champs parce qu'en utilisant le

descripteur MULTILINGUAL SYSTEMS, on n'a trouvé qu'une seule référence et avec le descripteur MULTILINGUAL INFORMATION SYSTEMS, on a trouvé deux références.

Pour le terme MULTILINGUAL on a eu 221 références qu'on a réduit à 67 en introduisant le champ LANGUE. La stratégie était:

- S.1 - MULTILINGUAL?
  - 221
- S.2 - MULTILINGUAL?/DE
  - 92
- S.3 - S2 AND LA=ENG
  - 60
- S.4 - S2 AND LA=FRENCH
  - 7
- S.5 - S.3 AND S.4
  - 67

Le tableau suivant donne une vue des bases consultées et des résultats obtenus.

SERVEUR	BASE DE DONNEES	REF. SORTIES
TEL.QUESTEL	PASCAL	43
DIALOG	LISA	67
TOTAL		110

**TABEAU 2**

### 3. Résultats et commentaires:

Le Tableau suivant donne le résultat final après le tri des références et l'exclusion des doublons:

BASE DE DONNEES	REF.	SORTIES	REF.	PERTIN.
PASCAL	43		13	
LISA	67		13	
TOTAL	110		26	

TABLEAU 3

Le Tableau 4 donne la répartition des références par année:

PASCAL		LISA	
ANNEE	REFERENCE	ANNEE	REFERENCE
		1988	3
1986	2	1986	1
1983	2	1983	1
1979	2	1982	3
1978	1	1981	1
1976	2	1980	1
1975	1	1979	1
1974	1	1978	1
1973	2	1977	1
TOTAL	13		13

TABLEAU 4

Malgré la mise à jour mensuelle des deux bases de données, on a constaté une plus grande actualité dans la base de données LISA. Par contre, les résumés trouvés sur LISA sont moins détaillés que ceux de PASCAL; ce qui nous amène à conclure que c'est une des raisons du manque d'actualité de PASCAL.

Pendant la description de la recherche manuelle, on a dit qu'on a trouvé au Laboratoire d'Informatique Documentaire un seul ouvrage; c'est le plus important car il s'agit des Actes du Troisième Congrès Européen sur les Systèmes et Réseaux Documentaires qui a comme thème: Franchir la barrière linguistique (Luxembourg, 1977).

Cet ouvrage recense 40 travaux présentés dans ce Congrès et donc, concernant directement le sujet en question.

Cependant, sur la base de données PASCAL, cet ouvrage était considéré en sa totalité, tout comme deux des travaux inclus; de la même façon, sur LISA, on a trouvé huit références des travaux du Congrès, analysés séparément.

On l'a considéré comme une seule référence et du fait qu'on a trouvé des doublons dans les bases de données, on ne les a pas pris en compte dans le total des références pertinentes.

Si on considère les résultats de la recherche manuelle (l'ensemble des références retenues dans les revues des sommaires et celles du Congrès) et les résultats de la recherche automatisée, on aura pour le total des références, un niveau raisonnable de pertinence et d'efficacité des réponses par rapport à la précision du sujet.

On a dans le Tableau 5 une comparaison entre la recherche manuelle et la recherche automatisée avec les références pertinentes sans les doublons:

Recherche manuelle	06
Recherche automatisée	36
Recherche automatisée (sans doublons)	26
TOTAL	32

**TABLEAU 5**

#### 4. Repérage des documents

Le repérage des documents se fait petit à petit avec l'aide des services de la Bibliothèque Universitaire, soit par le prêt inter-bibliothèques, soit par la commande de photocopie des originaux au CNRS et au BLDS.

On rencontre des problèmes pour accéder aux documents primaires, surtout ceux signalés sur LISA car les références sont incomplètes.

## 5. Bibliographie:

1.

ANDRES, M. Etat de l'utilisation du système d'information AGRIS dans le domaine de l'economie agricole, forestière et des industries alimentaires. Informatik 33(4):153-154, 1986.

2.

BERTRAND-GASTALDY, S. De quelques éléments à considérer avant de choisir un niveau d'analyse ou un langage documentaire. Documentation et Bibliothèques 32(1/2): 3-23, jan./fev. 1986.

3.

BONZI, S. Terminological consistency in abstracts and concrete disciplines. The Journal of Documentation 40(4):247-263, Dec. 1984.

4.

CURWEN, A.G. A decade of PRECIS, 1974-1984. Journal of Librarianship 17(4): 244-267, Oct. 1985.

5.

D'OLIER, J.H. Multilingualism in the scientific and technical documentation. International Forum of Information and Documentation 2(4): 20-24, 1977.

6.

DRUCKEREI, E.K. Euro-Arabic terminology data base. LISA, 1988.

7.

DUBOIS, C.P.R. Multilingual information systems: some criteria for the choice of specific techniques. Journal of Information Science 1(1): 5-12, Apr. 1979.

8.

DUCROT, J.M. Le système TITUS II. Association Nationale de Recherche Technique en Information et Documentation 4:3-40, 1973.

9.

DUCROT, J.M. Textile: TITUS. Seminaire International d'Information en Sciences de l'Information. Rabat, 1979. p.74-78.

10.

DUSOULIER, N. & IVANES, D.P. Problèmes théoriques et pratiques posés par les systèmes multilingues: attitude nationale et internationale. Premier Congrès National Français d'Information, Documentation et Communication. Paris, 1974. p. 249-256.

11

FIDEL, R. Writing abstracts for free-text search. The Journal of Documentation 42(1) Mar. 1986.

12.

GEBHARDT-CORDERO, H. et alii. Short European linguistic dictionary. Universitat Mannheim. LISA, 1988.

13.

HAINEBACH, R. European Legal Literature Information Service (ELLIS): a master guide to commentary on European Community law. LISA, 1988.

14.

HARGREAVES, P. On-line information retrieval from foreign languages data bases. (MSC Thesis) Loughborough University of Technology, 1980. 160p.

15.

HORSNELL, V. Report of a workshop on multilingual system. British Library Research Development Department. Report n. 5265. LISA, 1982.

16.

LAUZON, J. Les flux transfrontières de données. Documentation et Bibliothèques 31(3): 101-105, juil./sept., 1985.

17.

MILDNER, H. Le thésaurus international pour l'agriculture et exploitation forestière. Informatik 23(4): 31-33, 1976.

18.

NDIAYE, G. MINISIS: un progiciel pour l'exploitation multilingue des bases de données documentaires. Colloque International de Montreal (CAN.), 1983. p. 288-291.

19.

PADMANABHAN, V.S. & AMBA, K.N. Some factors to be considered in the compilation of a thesaurus for leather and allied industries. Seminary of Thesaurus Information Systems, Bangalore, 1975.

20.

PELISSIER, D. & ARTHUR, O. The multilingual evolution of PASCAL. Tenth International Online Information Meeting, London, 2-4 Dec. 1986. p.113-121.

21.

FIGUR, V.A. Systèmes de recherche de l'information multilingues: niveaux d'intégration et support linguistique. Nauch.-Tekh. Inform. 1:22-28, 1979.

22.

RISSE, K. Experiences dans la mise en oeuvre et le developpement de la coopération internationale bilatérale. Informatik 20(6):7-8, 1973.

23.

SAGER, J.C. et alii. Thesaurus integration in the social sciences: part I: comparison of thesauri. International Classification 8(3): 133-138, 1981.

24.

SAGER, J.C. et alii. Thesaurus integration in the social sciences: part II: stages towards integration. International Classification 9(1): 19-26, 1982.

25.

SAGER, J.C. et alii. Thesaurus integration in the social sciences: part III: guidelines for the integration of thesauri. International Classification 9(2):64-70, 1982.

26.

SCHOEN, J. Situation de l'exploitation de l'information dans les systèmes I&D de la République Fédérale Allemande. Kooperation in der Klassifikation, Frankfurt, 1978. p. 96-110.

27.

TAUPIN, P. L'automatique documentaire à l'OCDE. Documentaliste 20 (1):14-16, 1983.

28.

TENOPIR, C. Searching Harvard business review online lessons in searching a full text databases. Online Review 9(2): 71-78, Mar. 1985.

29.

TROISIEME CONGRES EUROPEEN SUR LES SYSTEMES ET RESEAUX DOCUMENTAIRES: franchir la barriere linguistique. Luxembourg, 3-6 mai 1977. Munchen: Verlag Dokumentation, 1977. 730p.

30.

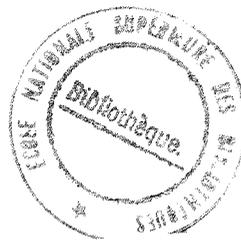
VAN SLYPE, M.G. Définition des caractéristiques essentielles des thesauri. Rapport final, I e II. Bruxelles, Bureau Marcel Van Dijk, 1976.

31.

WOLFF-TERROINE, M. SABIR: un logiciel de gestion de thésaurus multilingues. MIDIST Bulletin d'Information 2:8, Apr. 1983.

32.

ZINGEL, H.-J. Experiences with TITUS II. International Classification 5(1): 33-37, Mar. 1978.



BIBLIOTHEQUE DE L'ENSSIB



8412856