

Chapitre IV - La dimension discursive de l'indexation

L'objectif de ce chapitre est de définir le **processus** de l'indexation dans le cadre d'un modèle de fonctionnement de la langue, ou, plus précisément, dans le cadre d'une approche linguistique du lexique et de la référence.

Trouvant ses fondements dans un modèle linguistique du lexique et de la référence, notre approche de l'indexation en termes de discours ne sera pas elle-même de nature strictement linguistique. Notamment, la notion de discours, telle qu'elle sera traitée dans ce chapitre, se laissera approcher dans le cadre plus large d'une analyse sémiotique : sauf indication contraire, on entendra ici par discours un « ensemble d'énoncés et/ou de textes, **possédant une organisation** thématique, normative, structurale »¹. C'est principalement sous l'angle de l'organisation de textes que nous étudierons la dimension discursive de l'indexation : à ce titre, le discours documentaire se conçoit comme un espace d'organisation spécifique des documents obtenu par le biais d'« une transformation discursive réglée ».

Nous nous attacherons dans ce chapitre à dégager les **principes** qui sous-tendent les deux actes de référencement réalisés en indexation² :

- quels sont les principes qui dictent l'acte de discrétisation sur les sources ?
- quels sont les principes qui guident l'acte de stabilisation de l'espace documentaire ?

Ces deux types de principe seront analysés en termes de **stratégie** : on distinguera, d'une part, l'indexation en tant que stratégie d'exploration des sources et, d'autre part, l'indexation en tant que stratégie d'exposition des documents.

La dimension discursive de l'indexation est particulièrement absente des discours classiques : comme nous l'avons vu dans le chapitre II, c'est la dimension lexicale qui prédomine. De même, on s'interroge généralement peu sur les principes de sélection des sources en indexation, même si leur importance a pu être relevée³.

¹. Souchard 1989, p. 258 (c'est nous qui soulignons).

². Voir chapitre III.

³. Par Bertrand-Gastaldi 1986, par exemple, p. 4 : « Nous n'étudierons pas le sous-système de

Enfin, la notion de stratégie d'exposition des documents reste en général appréhendée par le seul biais des langages documentaires.

Il nous faut donc substituer, dans un premier temps, à la notion de langage documentaire, la notion de discours documentaire, de façon à aborder, dans un second temps, d'une part, la constitution de ce discours documentaire (stratégie d'exploration des sources) et, d'autre part, les conditions de son utilisation (stratégie d'exposition des documents).

Nous procéderons donc dans ce chapitre en trois étapes :

- nous tâcherons d'abord de définir ce que l'on peut entendre par discours documentaire, notion que l'on substituera à celle de langage documentaire. En distinguant langage et discours documentaires, nous cherchons à distinguer les mots (langage documentaire) des utilisations qui en sont faites (dans un espace de discours). Pour ce faire, nous examinerons comment l'approche linguistique de pratiques comme la terminologie et la vulgarisation scientifique a appréhendé cette question : sur ce point, les problématiques de l'indexation rejoignent celles d'autres pratiques ;
- nous nous intéresserons ensuite à l'espace de discours que construit l'indexation à partir des sources, en traitant la question suivante : sur quels principes l'indexation sélectionne-t-elle ses sources ? Nous nous appuierons sur les notions de « système-archive » et de « formation discursive » proposées par Foucault ;
- nous nous pencherons, pour finir, sur l'articulation que l'indexation met en place entre les trois espaces dans lesquels elle évolue : espace des auteurs (sources), espace des indexeurs (documents), espace des utilisateurs (objets de discours). Cette articulation sera pensée par le biais de la notion de « monde possible » proposée par Kripke.

collecte qui, pourtant, constitue le premier filtre important grâce à la sélection d'un nombre réduit de domaines du savoir, de publications, de supports, en fonction des objectifs poursuivis ».

I- Langage ou discours documentaire ?

Nous situerons, dans un premier temps, la question que nous posons (« langage ou discours documentaire ? ») dans le cadre plus large de la réflexion des pratiques sur elles-mêmes.

Nous aborderons ensuite l'approche linguistique dont ont fait récemment l'objet la terminologie et la vulgarisation scientifique, deux pratiques professionnelles qui nous paraissent pouvoir être fructueusement rapprochées des pratiques documentaires.

I.1- Notions de langage et de discours dans les pratiques : enjeux

Avancer la notion de discours documentaire peut sembler incongru dans la mesure où l'indexation ne crée pas de textes à proprement parler. Ce que nous cherchons à montrer en opposant langage documentaire à discours documentaire, c'est que l'indexation n'utilise pas un langage particulier, ce que laisse entendre la notion de langage documentaire, mais qu'elle fait une utilisation documentaire du langage (ou plutôt des productions du langage), ce que voudrait faire entendre la notion de discours documentaire.

Il nous semble tout à fait important de noter que, sur ce point, la pratique d'indexation n'est pas la seule pratique professionnelle à difficilement intégrer la dimension discursive dont elle relève. Nous avons retenu, à la fois pour exemples et pour guides, les pratiques de la terminologie et de la vulgarisation scientifique qui ont récemment connu un renouvellement de leurs problématiques dans un sens proche de celui que nous voudrions introduire dans cette recherche. Aussi bien la terminologie que la vulgarisation scientifique constituent des pratiques qui articulent des ensembles de mots (en l'occurrence des terminologies) et des ensembles de textes (textes spécialisés ou textes vulgarisés). La perception des pratiques par elles-mêmes tend à focaliser l'attention sur les mots eux-mêmes ; la perception linguistique des pratiques tend, elle, à distinguer les mots d'une part et leur rôle dans un type de discours d'autre part. C'est cette distinction que nous cherchons à mettre au jour dans la pratique d'indexation.

Sur ce point, il ne nous semble pas que les critiques courantes des langages documentaires permettent réellement de mettre à distance la dimension lexicale en indexation. On ne compte plus les condamnations des thésaurus et autres langages « autoritaires »⁴ et nombre de systèmes d'indexation automatique⁵ fondent leur

⁴. Voir, par exemple, Turner 1990.

⁵. Essentiellement ceux qui exploitent une liste inverse de chaînes de caractères. L'indexation est

argumentaire de vente sur le « texte intégral », autrement dit sur l'absence de langage contrôlé. Cependant, ces critiques ne remettent pas en cause la dimension lexicale de l'indexation ; elles discutent simplement la forme, plus ou moins artificielle et plus ou moins figée, des descripteurs.

Ce que nous voudrions montrer c'est que, si « lexique documentaire » il y a en indexation (sous forme de liste de descripteurs), il est à comprendre dans un cadre discursif, dans un cadre où les documents peuvent se constituer comme des discours qui s'interpellent *via* des mots. C'est dans ce cadre discursif que les descripteurs peuvent alors être approchés (Chapitre V).

I.2- Évolution des problématiques en terminologie : émergence de la notion d'usage professionnel de la langue

Il nous semble que la question « langage ou discours documentaire » s'éclaire comparée à la problématique « langue ou discours de spécialité » qui caractérise les débats en terminologie (I.2.1). Par ce nouvel éclairage, nous pourrions restituer au discours documentaire ses deux dimensions discursives - textuelle et énonciative - particulièrement peu visibles dans les discours normatifs (I.2.2).

I.2.1- Problématiques de la terminologie

La terminologie est avant tout une pratique professionnelle (historiquement un « art ») qui cherche à se constituer, depuis peu, en discipline scientifique spécifique, à la recherche de fondements théoriques, entre autres linguistiques⁶.

L'émergence d'une approche théorique s'est faite en terminologie sur la base de reformulations importantes. L'enjeu a notamment consisté à souligner que la terminologie ne pouvait se concevoir sous l'angle du seul lexique, comme « un lexique spécialisé » ou encore comme une « langue spécialisée », mais qu'elle

entendue dans ce cas dans une acception informatique, qui repose sur la notion informatique d'« index » : « L'index d'une banque de données texte intégral est le plus souvent représenté sous la forme d'un fichier inverse. Dans un fichier inverse, un enregistrement est créé pour chaque terme apparaissant dans la collection des documents à indexer. À chacun de ces termes est associée une liste de couples de la forme (identificateur de fichier, position dans le fichier). Chaque élément de cette liste correspond à une occurrence du terme dans la collection des documents. Si le mot *avion* apparaît dans le fichier 12 à la position 16 et dans le fichier 15 à la position 18, le fichier inverse contient une entrée de la forme : *avion* (12 16) (15 18) ». Rôle 1993, p. 137.

⁶. Un aperçu historique se trouve dans Rey 1992 ; de façon plus précise, les prémices de la réflexion terminologique sont identifiées dans Mustafa-Elhadi 1989. Lerat 1995 propose une synthèse des approches théoriques.

devait se comprendre sous l'angle du discours⁷. Dès lors, c'est la notion même de terme* qui s'est trouvée redéfinie. Elle ne correspond plus à un type de mot particulier, caractérisé par sa monoréférentialité (en opposition aux mots non spécialisés qui peuvent, eux, être polyréférentiels) ; elle correspond plutôt à un usage particulier d'un certain type de mots. La notion de terme se laisse alors capter à plusieurs niveaux, à la fois à un niveau linguistique (en termes de propriétés) et à un niveau extra-linguistique (en termes d'usages sociaux, professionnels).

En tant que pratique définie par des normes, la terminologie⁸ distingue les notions de « langues de spécialité » et de « terminologie »⁹, celle-ci étant intégrée à celle-là :

On entend par langue de spécialité « un sous-système linguistique qui utilise une terminologie et d'autres moyens linguistiques et qui vise la non-ambiguïté de la communication dans un domaine particulier ».
ISO 1087, 1990 cité in Lerat, 1995, p. 17.

D'emblée, notamment d'un point de vue linguistique, la notion de « langue de spécialité » apparaît problématique, dans la mesure où elle ne saurait être assimilée à un type de langue, à un dialecte : la langue est, en terminologie, présente telle qu'en elle-même. Elle est, en revanche, comme dans l'indexation, utilisée dans une perspective particulière, professionnelle :

Le français de l'automobile est l'usage du français pour rendre compte de connaissances en matière d'automobile. [...] La langue spécialisée est d'abord une langue en situation d'emploi professionnel (une « langue en spécialité » comme dit l'école de Prague). C'est la langue elle-même (comme système autonome) mais au service d'une fonction majeure : la transmission des connaissances.
Lerat, 1995, p. 21 (c'est nous qui soulignons).

Ce changement de perspective, de « langue spécialisée » à « usage spécialisé », conduit les théoriciens à privilégier la notion de « discours de spécialité », rejetant comme impropre la dénomination de « langue de spécialité »¹⁰.

Avec la notion de discours de spécialité se repositionne celle de « terminologie » (en tant que liste de termes). Elle n'est plus dans un rapport hiérarchique d'inclusion (comme l'indique la norme ISO), mais dans un rapport dynamique de définition

7. Voir Le Guern 1989 notamment.

8. L'ISO définit la terminologie comme l'« étude scientifique des notions et des termes en usage dans les langues de spécialités », ISO 1087 (1990).

9. « Terminologie » est à entendre ici « comme un ensemble d'expressions dénommant dans une langue naturelle des notions relevant d'un domaine de connaissances fortement thématisé », Lerat 1995, p. 20.

10. Voir Rey par exemple : « Il n'y a pas à proprement parler de “langue” mais des “vocabulaires”, des “usagers” et des “discours de spécialité” », ou encore Quémada : « La linguistique descriptive [...] condamne les désignations de “langue” technique et scientifique qui sont également impropres ». Citations extraites de Kocourek 1991a, p. 15.

avec les textes spécialisés :

Les textes savants saisissent et expriment le contenu savant, dont les unités sémantiques dominantes sont les termes. Le terme est une unité lexicale - ou son acception - définie dans les textes savants, où apparaissent ses occurrences intégrées dans le tissu du texte. [...] Les termes ne sont pas seulement des éléments du système, mais des occurrences dans les textes savants.

Kocourek, 1991b, respectivement p. 71 et p. 72.

Mais les textes spécialisés ne se contentent pas de définir les termes ; ils les constituent également en tant que tels, dans la mesure où :

Ce n'est que l'emploi répété du terme dans les textes différents qui mène, au fur et à mesure, à son implantation, à sa vraie terminologisation, à son passage du texte au système. Ce sont les textes spécialisés qui décident du statut systémique d'un néologisme.

Kocourek, 1991b, p. 73.

Ainsi l'existence des termes apparaît-elle complètement déterminée par des textes dont cependant le caractère « spécialisé », « savant », est lui-même fortement déterminé par la présence des termes ...

La dimension textuelle ne permet donc pas de constituer à elle seule un « discours spécialisé » ; intervient également une dimension énonciative, celle des locuteurs :

En fait, la dichotomie langue générale/langues de spécialité se fait à partir de l'opposition expérience partagée par l'ensemble des locuteurs/expérience partagée par des sous-ensembles de locuteurs.

Portelance, 1989, p. 402.

Ainsi les notions de discours spécialisé et de terminologie semblent-elles se constituer autour de l'idée d'**espace d'expérience partagée**.

Rien de moins hypothétique, bien entendu, qu'un découpage en cercles, plus ou moins larges, de locuteurs ; c'est pourtant bien l'idée d'une convention entre locuteurs qui légitime et fonde une utilisation partielle et spécialisée de la langue en terminologie.

C'est ainsi que la notion d'espace discursif a été, en terminologie, peu à peu explicitement posée comme centrale par les chercheurs du domaine¹¹. C'est elle qui permet de concevoir la notion de concept* à laquelle se réfère la pratique terminologique : il ne s'agit plus de concepts liés aux mots eux-mêmes mais de concepts liés aux objets scientifiques manipulés par un nombre restreint de locuteurs. Comme l'indique Mortureux, « ce type de discours [le discours spécialisé] s'échange en général à l'intérieur d'un cercle de spécialistes, et par conséquent les questions que peut soulever la réception des termes sont intimement

11. Voir par exemple, Boutayeb 1993, p. 14 : « Dans les unités terminologiques, l'expression (c'est-à-dire la forme) résulte d'une convention qui est le résultat de l'accord des spécialistes d'un domaine scientifique donné. Le terme s'impose par une pratique unifiée dans un milieu d'experts ».

liées aux concepts et relèvent plus de la démarche scientifique que de la compréhension strictement linguistique »¹².

Une approche linguistique de la pratique terminologique permet de mettre en valeur les deux facettes de la dimension discursive des terminologies (dimension textuelle et dimension énonciative), rendant délicate toute approche qui considérerait les termes comme des types de mot particuliers :

- d'une part, elle montre que les termes des terminologies sont indissociablement liés aux discours qui les instituent ;
- d'autre part, elle souligne que les termes ne sont pertinents¹³ que dans le cadre restreint d'une pratique technique ou scientifique donnée.

Si la terminologie vise « la non-ambiguïté de la communication », comme l'indique l'ISO, ce ne peut être que dans un cadre doublement restreint et par un ensemble de textes (notion de discours spécialisé) et par un ensemble de locuteurs (notion de cercles étroits de spécialistes). Créant un espace de textes et dessinant un espace de locuteurs, la terminologie se conçoit alors comme une pratique qui spécifie un usage « spécialisé » de la langue et non plus comme une pratique qui construit des « langages de spécialité » ou des « langues spécialisées ». C'est ainsi que la pratique terminologique consiste essentiellement¹⁴ à construire des corpus de textes « spécialisés » et à établir, à partir d'eux, le mode d'emploi de certains mots, les termes, et ce, à l'intention non du « grand public » mais « des locuteurs spécialistes du domaine d'origine ou d'un domaine connexe »¹⁵.

En quoi les problématiques de la terminologie et le regard croisé, pratique et théorique, dont elles ont fait l'objet peuvent-elles nous permettre d'approcher la dimension discursive de l'indexation ?

I.2.2- Rapprochement de deux disciplines

On notera tout d'abord que le rapprochement entre pratique documentaire et pratique terminologique n'est pas nouveau : il est même en plein essor.

¹². Mortureux 1995, p. 23.

¹³. Les termes ont pour fonction de réaliser une désignation référentielle univoque (un terme = un concept ; un concept = un terme). On peut dire que leur pertinence est liée à la réalisation de cette fonction.

¹⁴. Mais pas uniquement, voir sur ce point, dans le glossaire, l'entrée « analyse terminologique ».

¹⁵. Mortureux 1995, p. 23.

Michel Le Guern et les membres du groupe SYDO¹⁶ ont, dès les années 80, suscité l'attention des professionnels : les propriétés du terme comme unité de discours méritent d'être rapprochées des fonctionnalités attendues du descripteur.

Ce n'est que dix ans après, et dans des termes sensiblement différents, que la pratique terminologique est rapprochée de la pratique d'indexation par les professionnels eux-mêmes. Les années 1990 ont vu en effet se multiplier les rencontres « terminologie/documentation »¹⁷. Parallèlement, à la même époque, les concepteurs de systèmes d'indexation automatique ou automatisée¹⁸ renouvellent leur argument de vente : l'utilisation des terminologies pour l'indexation et la recherche documentaires permet d'assurer des taux de rappel* et de précision* bien supérieurs à ceux obtenus par l'utilisation des seuls thésaurus¹⁹. Ce qui retient l'attention, semble-t-il, des professionnels de l'information, c'est le projet de la terminologie d'établir une « communication non ambiguë » *via* les termes d'une terminologie. Les professionnels fondent leur espoir sur cette caractéristique du terme sans toujours observer que celle-ci suppose à la fois un domaine restreint de textes et un cercle circonscrit de locuteurs. Dans une telle perspective d'utilisation des terminologies en indexation, il s'agit toujours de recourir à un lexique, ici constitué de termes, pour réaliser l'indexation : on ne sort pas à proprement parler du domaine lexical et on ne prend pas en compte la dimension discursive des termes²⁰.

Si la pratique de l'indexation peut être, utilement, nous semble-t-il, rapprochée de la pratique terminologique, elle ne peut cependant complètement s'y fondre. Les raisons sont nombreuses qui peuvent permettre d'argumenter ce point. Nous n'en retiendrons qu'une dans cette recherche : alors que la pratique terminologique vise à établir une communication entre spécialistes, une communication « spécialisée », la pratique d'indexation vise à établir une communication plus large entre non-spécialistes (ou du moins pas nécessairement spécialistes) au sein de cercles de

16. Pour les textes fondateurs, voir Le Guern 1984, Bouché 1989, Metzger 1988.

17. Par exemple, en 1993, le département *Formation continue* de l'Institut d'Études Politiques de Paris organise un séminaire de trois jours, intitulé « Terminologie et documentation » ; l'expérience est renouvelée depuis tous les ans. La Direction de l'information scientifique et technique du ministère de l'Enseignement supérieur et de la recherche met l'accent sur ce thème en 1994 en organisant la première journée « Terminologie et Information ». Dernier exemple en date : l'ADBS (Association des professionnels de l'information et de la documentation) a organisé, dans sa journée d'étude du 19 décembre 1996 consacrée aux « Outils linguistiques et [aux] nouvelles technologies », une table ronde sur les synergies entre terminologie et thésaurus.

18. Dans le sens donné en note 5.

19. Voir par exemple le colloque RIAO (Recherche d'information assistée par ordinateur) de 1994 qui relève dans son compte rendu, p. 13 : « Les descriptions formelles du terme que nous fournissent les terminologues peuvent être utilisées pour automatiser, avec beaucoup de succès, la reconnaissance des descripteurs » ; des systèmes automatisés sont présentés dans le même volume.

20. Nous proposons dans le chapitre V une tout autre approche des termes en indexation qui reprend les présupposés établis par Le Guern et les membres de l'équipe SYDO.

locuteurs appartenant à des horizons divers : on parlera alors de communication « vulgarisante ».

Si, à un niveau qu'il conviendra de préciser, les deux pratiques terminologique et documentaire peuvent être rapprochées, il convient d'abord de montrer qu'elles mettent en jeu des problématiques similaires.

L'approche linguistique de la terminologie a permis de passer d'une perception de la terminologie comme « langue de spécialité » à une appréhension en termes de « langue en spécialité », une langue en situation d'usage professionnel. Un tel déplacement nous semble pouvoir être réalisé en indexation. On cherchera à montrer que la notion de langage documentaire est inadéquate et qu'il faut lui substituer la notion de discours documentaire, au sens d'usage documentaire des mots. Pour cela, nous nous appuyons sur les deux aspects de la dimension discursive mise en valeur par l'approche linguistique de la terminologie : comment se manifestent en indexation la dimension textuelle et la dimension énonciative des descripteurs ?

A- Dimension textuelle des descripteurs

De la même façon que les termes sont créés par et dans les discours spécialisés, on peut dire que les descripteurs sont créés par et dans des discours que l'on qualifiera de documentaires²¹.

Le discours normatif de l'indexation, qui met en avant la notion de langage documentaire, ne dit rien de l'origine textuelle des mots retenus pour figurer dans ce langage au titre de descripteurs²². Or un langage documentaire est toujours constitué de mots extraits d'un corpus de textes, de « mots de discours ». Nous illustrerons cet aspect des langages documentaires par quelques exemples concernant la constitution (A1) et la mise à jour (A2) de thésaurus.

A1- Les normes²³ et les manuels²⁴ indiquent deux méthodes pour collecter les termes²⁵ d'un thésaurus, qui toutes deux consistent en une extraction de mots du discours.

Dans la première méthode, dite inductive (ou synthétique), les termes sont extraits de référentiels existants (dictionnaires spécialisés, nomenclatures, terminologies,

21. Dans le sens de discours que l'on a précédemment donné : discours comme espace spécifique d'organisation textuelle.

22. Suzanne Bertrand-Gastaldi [1993, p. 147] note, sur ce point, que si « thésaurus et plan de classification sont tous deux le résultat d'une transposition d'énoncés antérieurs ou synchroniques, [le] corpus qui a servi à leur constitution est difficilement cernable, il est flou, non déclaré ».

23. Notamment norme Z 47-100 (1981).

24. Par exemple, Chaumier 1978.

25. L'approche normative [norme Z 47-100 (1981)] ne définit pas explicitement ce qu'elle entend par « terme » ; dans certains cas, le terme est bien celui de la terminologie ; dans d'autres contextes, l'emploi paraît plus approximatif.

tables de matière, classifications, etc.) ; dans la seconde méthode, dite déductive (ou encore analytique), les termes sont extraits d'un corpus de documents (de 1 000 à 5 000 documents). Généralement, il est conseillé d'employer les deux méthodes et de recourir à des « spécialistes » pour déterminer les termes importants d'un « domaine ». Sur ce plan, on le voit, la collecte des descripteurs se différencie peu d'une collecte de termes que réaliserait un terminologue. Cependant, si un terminologue a pour charge de référencer avec précision la source dont il extrait un terme, le compilateur d'un thésaurus, lui, s'affranchit plus radicalement de l'origine textuelle des termes qu'il retient.

A2- C'est sur le même mode d'extraction de mots de discours que s'effectue la mise à jour des thésaurus. Un terme est repéré dans la littérature comme susceptible de désigner un « concept » nouveau ; il est alors constitué comme candidat-descripteur généralement jusqu'à ce que le nombre de ses apparitions dans les documents justifie son insertion dans le thésaurus. Un terme repéré dans les documents à indexer peut être « préféré » à un descripteur utilisé jusqu'alors : les deux termes peuvent être mis en relation d'équivalence.

Dans le cadre de l'expérience que nous avons réalisée²⁶, l'un des organismes documentaires a repéré un candidat-descripteur dans la source du *Monde* à indexer que nous avons retenue : il s'agit du terme « redressement judiciaire », employé à plusieurs reprises dans l'article consacré au verdict rendu par le Tribunal de commerce sur les entreprises dirigées par Bernard Tapie²⁷. Le thésaurus utilisé par l'organisme documentaire comportait le terme « règlement judiciaire » ; ce terme est apparu aux indexeurs à la fois trop imprécis et trop en décalage avec le vocabulaire utilisé par les médias pour rendre compte de la longue série d'épisodes concernant le traitement judiciaire des sociétés de Bernard Tapie. Le terme « redressement judiciaire » a donc été constitué comme descripteur, le terme « règlement judiciaire » ayant été, lui, maintenu à ses côtés comme « synonyme ». Dans certains cas, la mise à jour de thésaurus, par adjonction comme par suppression de descripteurs, peut amener à une réindexation de tous les documents indexés par les « anciens » descripteurs. On voit jusqu'où peut aller, dans les pratiques d'indexation, le brouillage de l'origine textuelle des descripteurs.

Sur ce point, nous nous attacherons à montrer que tout ce qui fait le poids « informatif » d'un terme tient précisément au discours auquel il renvoie : la différence introduite dans un thésaurus entre « règlement judiciaire » et

²⁶. Voir Annexe 2.

²⁷. *Le Monde* du 1/12/1994, p. 24.

« redressement judiciaire » n'est significative que si l'on connaît les emplois qui les distinguent, les dates, les lieux, les contextes politiques ou sociaux dans lesquels ils apparaissent.

Le rôle des discours dans la constitution d'un lexique documentaire, tout comme le statut originel des descripteurs comme mots de discours, sont complètement opacifiés par la notion de langage documentaire compris comme « langage artificiel ». Or il nous paraît essentiel de faire apparaître la dimension textuelle des descripteurs : c'est elle qui permet aux descripteurs de pouvoir fonctionner comme objets de discours ; c'est elle qui rend indispensable la réflexion sur la notion de corpus et de collection en indexation. C'est en ce sens que l'indexation pourra se comprendre comme une stratégie d'exploration des sources (§ II, ci-après).

B- Dimension énonciative en indexation

Tout comme la dimension textuelle, la dimension énonciative reste totalement implicite dans le discours classique.

Si, comme l'indiquent normes et manuels, les termes retenus dans la littérature pour figurer en tant que descripteurs doivent être déterminés par un « spécialiste du domaine », que deviennent ces termes lorsqu'ils sont extraits de leur contexte et utilisés dans le cadre de cercles larges de locuteurs ? Comment la « non-ambiguïté » qui les définit comme termes peut-elle encore rester active ?

Le discours normatif n'appréhende pas cet aspect de la question, tant il semble évident que la stabilité des termes acquise dans le cercle étroit des indexeurs au contact des textes ne peut que tout naturellement se maintenir jusqu'aux utilisateurs. Ni la question de l'utilisation documentaire des termes ni celle de leur mode de communication ne sont posées, parce que, nous semble-t-il, n'est pas posée la dimension énonciative de l'indexation.

En considérant l'indexation comme espace de discours constitué de ses propres objets (les documents) et de ses propres locuteurs (les indexeurs), on parvient à mettre au jour les enjeux de l'indexation : il s'agit moins de partager les mêmes mots que le même univers, le même espace d'utilisation des termes ; c'est par l'éclairage apporté par l'approche théorique de la pratique terminologique que l'on peut déterminer un tel enjeu.

Reste la question du devenir des terminologies utilisées dans des cadres où elles ne sont *a priori* plus valides. Cette question concerne tout autant l'approche théorique de la terminologie que l'approche théorique d'autres pratiques, qui, comme l'indexation ou la vulgarisation scientifique, utilisent des termes :

Si les termes apparaissent, sont formés et « institués » dans et par les discours spécialisés (scientifiques ou techniques), ils n'en figurent pas moins dans quantité d'autres discours, et c'est, en règle générale, leur fonctionnement dans la communication étendue à des non-spécialistes (ou des « moins spécialistes ») qui pose avec le plus d'acuité les problèmes de leur interprétation, de leur fonction, de leur bien-fondé.

Mortureux, 1995, p. 22.

Le recours à l'évolution des problématiques en terminologie nous a permis d'approcher un aspect du discours documentaire en particulier : la dimension textuelle des descripteurs.

En terminologie, cette dimension apparaît comme ce qui constitue les termes dans leur usage et leur fonction professionnels : établir une communication spécialisée dans le cadre d'un cercle étroit de locuteurs. Quel rôle jouent, en indexation, les termes ? De quel type de communication relèvent les descripteurs ? Sur ce point, l'utilisation des terminologies dans les discours de vulgarisation scientifique apparaîtra particulièrement éclairante : elle permettra de préciser la dimension énonciative de l'indexation, que l'approche théorique de la terminologie avait permis de faire apparaître.

I.3- Évolution des problématiques en vulgarisation scientifique : émergence de la notion de discours stratégique

Tout comme la pratique terminologique, la pratique du discours de vulgarisation a récemment bénéficié d'un effort de recherche, qui a fait apparaître une distinction entre le lexique utilisé (des termes « spécialisés ») et l'objectif de communication propre à ce type de discours.

L'enjeu, pour les théoriciens, est de montrer comment, dans le discours de vulgarisation, les termes peuvent rester des termes : c'est par la notion de « discours de reformulation » qu'est décrit le contact maintenu entre discours de vulgarisation et discours spécialisés (I.3.1). Comme dans le paragraphe précédent, nous nous essayerons à un rapprochement entre disciplines : le rapprochement proposé ici est plus inédit et moins évident que précédemment puisque, contrairement à la vulgarisation scientifique, l'indexation documentaire, proche en cela de la terminologie, ne produit pas de discours à proprement parler. Le rapprochement nous semble néanmoins pouvoir être effectué sur la base des objectifs de communication, semblables, entre indexation et vulgarisation scientifique (I.3.2).

I.3.1- Problématiques de l'analyse du discours de vulgarisation scientifique

On essaiera ici de synthétiser l'essentiel des problématiques de l'analyse du discours de vulgarisation en mettant l'accent sur le déplacement qu'ont opéré les théoriciens²⁸ : l'analyse du discours de vulgarisation est passée d'une étude du lexique à une étude du discours²⁹. Chemin faisant, les analystes se sont intéressés moins aux termes pour eux-mêmes qu'aux stratégies de communication dans lesquelles ils étaient utilisés.

En reprenant la définition qu'en propose Mortureux, on peut entendre par discours de vulgarisation :

Un ensemble d'énoncés qui ont en commun d'assurer la diffusion de connaissances en dehors des cercles étroits de spécialistes qui les produisent.
Mortureux, 1989, p. 43.

Les « vulgarisateurs », comme les appelle Jacobi, s'interrogent peu sur leur pratique³⁰ qui bénéficie d'une longue tradition³¹. Reste que leur pratique repose sur un « savoir spontané » qu'il peut être utile de mettre au jour notamment afin de constituer le discours de vulgarisation comme objet d'étude. C'est à cette entreprise que se consacre Jacobi [1987] dont nous ne retiendrons qu'un aspect de la critique de la perception des pratiques par elles-mêmes.

A- Le mythe du « troisième homme » : la traduisibilité de la science
Parmi l'ensemble des mythes ou postulats que Jacobi met au jour en examinant les représentations que les praticiens se font de leur profession, nous retiendrons celui qu'il nomme le « mythe du troisième homme », mythe qui repose à la fois sur une certaine conception de la science et sur une certaine conception de la langue : « les spécialistes emploieraient volontiers un jargon obscur ou incompréhensible par les tiers à seule fin de tenir hors de portée un savoir sur lequel se fondent leur compétence et leur autorité »³². Autrement dit, à aucun moment, la science n'est

28. Selon les chercheurs du domaine, il n'existe pas véritablement, du moins à ce jour, de « théorie de la vulgarisation », mais plutôt un « ensemble de travaux convergents qui définissent un champ », Jacobi 1988, p. 12.

29. Conduite dans le cadre plus général de l'analyse de discours telle que Pêcheux a pu la définir (*supra*), Mortureux et Petit 1989 par exemple.

30. Cf. Jacobi 1988, p. 13 : « La pratique vulgarisatrice ne fait pas, ou très peu, l'objet d'une distanciation de la part des vulgarisateurs eux-mêmes. C'est une pratique qui semble se suffire à elle-même sans autre justification que sa propre production. Dans l'ensemble, la pratique vulgarisatrice est une pratique qui ne se pense pas ».

31. On s'accorde généralement à voir dans l'ouvrage de Fontenelle *Entretiens sur la pluralité des mondes* (1686) le premier exemple de texte de vulgarisation. Pour un aperçu historique, on peut se référer à Lazslo 1993, Jeanneret 1994 ou encore Raichvarg et Jacques 1991.

32. Jacobi 1988, p. 20.

appréhendée par les praticiens sous l'angle de sa complexité conceptuelle (la seule complexité de la science est d'ordre linguistique) ; à aucun moment non plus, la langue n'est envisagée sous l'angle de sa complexité linguistique (la seule complexité de la langue vient des rapports de pouvoir ou de domination qu'elle instituerait par le biais des mots). On remarquera enfin que la complexité des processus de communication dans lesquels la langue intervient n'est pas non plus abordée.

Sur la base de ce postulat d'une « traduisibilité de la science », le vulgarisateur se pose sous la figure du troisième homme, indispensable sauveur, entre homme de la science et homme de la rue³³. Dès lors, l'essentiel de sa pratique repose sur une traduction lexicale d'un registre de langue à l'autre. C'est sur ce point qu'une analyse linguistique du discours de vulgarisation scientifique peut montrer que cette perception de la pratique vulgarisatrice est biaisée :

On a souvent comparé le travail du vulgarisateur à un travail de traducteur : il réécrirait la science, exposée par les chercheurs dans un jargon incompréhensible, avec les mots de tous les jours. Cette image, pourtant séduisante, est inexacte. Si la science aime à fabriquer des mots nouveaux, c'est que les mots usés et polysémiques de notre langue commune semblent comme incapables de dire la science. Et en remplaçant les termes scientifiques par des synonymes approximatifs, on ne peut que déformer, transformer, réduire, caricaturer, bref dénaturer la science.
Jacobi, 1993, p. 81.

Une analyse linguistique des discours de vulgarisation scientifique montre que, contrairement à toute attente, les vulgarisateurs recourent systématiquement, dans leurs discours, aux terminologies spécialisées. Cette présence récurrente des termes spécialisés passe souvent inaperçue, prise qu'elle est, dans le discours qu'ils construisent, par les stratégies de communication qu'ils déploient³⁴.

Dès lors que la vulgarisation scientifique ne se pense plus dans les termes d'une traduction mot à mot, son enjeu se laisse redéfinir :

Vulgariser est une entreprise qui se situe au cœur d'une contradiction : comme le scripteur se propose de faire connaître le sens des notions et des concepts spécialisés construits par les sciences, il est contraint d'utiliser les termes et lexies des langues de spécialité ; mais, en employant dans son texte des termes spécialisés, il redoute - à juste titre - que les lecteurs ne puissent en comprendre le sens ; pour prévenir les difficultés d'accès au sens des destinataires, le scripteur recourt à une série de mécanismes, de type métalinguistique le plus souvent, qui lui permettent de mettre en relation les termes scientifiques avec les mots connus de la langue commune.
Jacobi, 1993, p. 81.

³³ *Ibid.*, p. 24 : « Ils [les vulgarisateurs] se perçoivent comme les intermédiaires naturels et essentiels chargés de combler l'écart entre les scientifiques et le grand public afin de rétablir une communication interrompue. Ils se présentent comme les artisans d'une possible révolution du savoir qui vise un partage plus équitable entre tous les acteurs sociaux ».

³⁴ *Cf.* Jacobi 1993, p. 77 : « Cette fonction [reformulation discursive sur la base des termes] est à ce point intégrée au discours que ces opérations, à la différence du dictionnaire, qui les affiche ou les exhibe, sont comme cachées dans le texte ».

B- Réintégrer la dimension discursive de la vulgarisation scientifique

Le courant de l'analyse linguistique³⁵ du discours de vulgarisation³⁶ s'est attaché à redéfinir le discours de vulgarisation scientifique dans des termes qui fassent apparaître le type de transformation que ce genre de discours opère : le discours de vulgarisation a ainsi pu être défini comme une transformation discursive de textes sources³⁷.

L'analyste s'attache dès lors à examiner, dans le discours second (le discours de vulgarisation), les traces des discours premiers (les textes scientifiques)³⁸. La présence des textes sources se signale par la présence de termes spécialisés dans le discours second. Comment identifier les termes puisqu'ils sont extraits du contexte de production qui les constitue comme termes ? Comment identifier les sources qui ne s'expriment, dans les discours seconds, que sous la forme des termes ?

Sans entrer dans le détail de la méthode d'analyse adoptée, on dira simplement que l'observation porte sur les stratégies discursives utilisées par les vulgarisateurs. Ces stratégies sont de deux types (définitionnel et désignationnel³⁹), et se laissent identifier par une série d'indices de nature diverse, typographique (emploi des guillemets par exemple), linguistique (séquences paraphrastiques par exemple), etc. Les termes sont identifiés *via* le repérage de ces stratégies discursives menées sur un corpus de textes de vulgarisation. La méthode de l'analyse du discours de vulgarisation relève en effet de l'analyse interdiscursive, reposant sur l'hypothèse que le discours de vulgarisation scientifique s'inscrit dans un réseau de textes, défini par une « formation discursive »⁴⁰. Cette analyse de l'emploi de termes spécialisés dans plusieurs discours de vulgarisation, comparé à l'emploi de ces mêmes termes dans plusieurs discours spécialisés, montre que le terme « scientifique » employé par les vulgarisateurs fonctionne différemment dans les textes scientifiques et dans les textes vulgarisés : dans les premiers, il révèle une certaine monoréférentialité (référence quasi constante au même concept) ; dans le

³⁵. Nous n'aborderons pas ici le courant sociologique de l'analyse du discours vulgarisateur qui examine les présupposés idéologiques d'une telle activité ; sur ce point, on peut, par exemple, se reporter à Boltanski et Maldidier [1977].

³⁶. Représenté entre autres par Mortureux [1983] et Jacobi [1984] pour les textes fondateurs.

³⁷. « On sait que la diffusion des connaissances s'effectue notamment à travers une activité discursive de reformulation des sources dites " primaires " », Mortureux 1993, p. 3.

³⁸. Cf. Mortureux 1988, p. 119 : les discours de vulgarisation « sont des discours " seconds ", dont la production, le fonctionnement et la légitimité renvoient à des « discours primaires » (dits parfois ésotériques), qui sont les publications par lesquelles les chercheurs exposent à leurs pairs les résultats de leurs travaux. Pour un linguiste, c'est là une première propriété de la vulgarisation scientifique, engageant à examiner, dans les énoncés réalisés, les modalités de cette secondarité, les manifestations de cette référence aux discours primaires, les traces de la réénonciation qui les a produits ».

³⁹. Nous y revenons au paragraphe II de ce chapitre.

⁴⁰. Jacobi 1988, p. 38-39. Pour une présentation de la notion empruntée à Foucault, voir le § II.1 de ce chapitre.

second, il révèle une grande plasticité référentielle (variabilité des catégorisations référentielles).

Le rôle des termes en vulgarisation n'est donc pas comparable à celui qu'ils tiennent en terminologie : les termes, dans le texte de vulgarisation, sont là pour focaliser l'attention du lecteur sur les passages explicatifs ou définitoires, et non pour permettre au lecteur d'établir une relation référentielle stable⁴¹. Ils tiennent un rôle de « termes pivots » : comme isolés du discours par les stratégies discursives dont ils sont l'objet, ils apparaissent comme des citations, des extraits d'autres textes sur lesquels le texte présent s'appuierait pour définir, pour reformuler le savoir scientifique.

Ce que montre l'utilisation des termes dans les textes de vulgarisation, par opposition à l'utilisation des termes dans les discours spécialisés, c'est que le « savoir [y] est décontextualisé »⁴². Comme a pu également le souligner Grize, les sources scientifiques sont, dans le discours de vulgarisation, détournées de leur usage initial⁴³ ; et c'est grâce à ce détournement, à cette décontextualisation, que le « savoir » peut être transmis. En effet, par les déplacements référentiels opérés à partir des « termes pivots », le vulgarisateur déploie une série de désignations multiples qui permet des saisies et des ressaisies, diverses, des concepts scientifiques, par les lecteurs non spécialistes : ce n'est donc pas exactement le « même » savoir qui est diffusé entre spécialistes et à destination de non-spécialistes mais le lien est maintenu par le biais du terme, porteur d'un ancrage discursif dans les textes scientifiques.

Cette présentation de l'évolution des problématiques en vulgarisation scientifique, d'une approche centrée sur la traduction lexicale à une approche révélant des faits de discours et d'interdiscours, nous paraît de nature à préciser les enjeux de la dimension discursive en indexation.

⁴¹. Cf. Mortureux 1988, p. 145 : « La vulgarisation scientifique contemporaine ne fournit pas un apprentissage de la terminologie : les termes qu'elle cite sont généralement isolés des ensembles conceptuels et formels qui les structurent [...] et leur paraphrase n'attire guère l'attention sur cette systématisme des terminologies. [...] Le discours de vulgarisation scientifique se borne à faire référence, allusion, en mentionnant un certain nombre de termes. Là encore, on peut légitimement parler d'information plus que de formation ».

⁴². Jacobi 1988, p. 22.

⁴³. Pour Grize, le propos de la vulgarisation consiste à « communiquer des savoirs à d'autres fins que leur mise en pratique », cité in Jacobi 1987, p. 8.

I.3.2- Rapprochement de deux disciplines

Le rapprochement entre indexation et vulgarisation scientifique, s'il semble être autorisé par la conception de la pratique vulgarisatrice que propose Jacobi⁴⁴, n'a pas été, à notre connaissance, établi.

On en trouve néanmoins, nous semble-t-il, une incitation dans une contribution que Sylvain Auroux présente dans un ouvrage collectif consacré aux sciences en bibliothèque⁴⁵. Il émet l'hypothèse que l'accès au savoir dont la documentation a la charge doit s'inspirer des méthodes employées par la vulgarisation scientifique puisque l'enjeu consiste à éviter une « distorsion » entre les représentations de la science et les représentations communes : « ce dont nous avons besoin, c'est inventer et installer dans notre familiarité une forme de représentation admissible des connaissances qui soit en cohérence avec la pratique de notre vie quotidienne »⁴⁶. Auroux illustre son propos ; le bibliothécaire apparaît comme celui qui doit établir une connexion entre un long article sur les particules et deux lignes sur ce sujet dans une encyclopédie : entre les deux, il y a, dit-il, « une chaîne de diffusion de savoir » à établir⁴⁷.

Notons que le rapprochement de l'indexation et de la vulgarisation tend à redéfinir la finalité de l'indexation, voire plus largement, celle de la documentation : il s'agirait non plus de fournir des informations mais de donner accès aux savoirs.

La distinction entre information et connaissance constitue le cœur de débats importants dans le champ des sciences de l'information et de la communication⁴⁸. L'angle d'approche que nous avons retenu ne permet pas réellement de donner corps à ces distinctions. En effet, nous avons jusqu'à présent essayé de préciser ce que, dans le cadre d'un modèle linguistique, pouvait être l'information telle que la norme d'indexation la présente : « extraite » d'un texte, « traduite » dans un langage documentaire, « trouvée » par un utilisateur. En dégagant cette formulation des présupposés réalistes dans lesquels elle semblait prise, nous avons

44. Jacobi [1987, p. 31] tient que la vulgarisation scientifique n'est pas « une pratique spécifique » ; pour lui, « la vulgarisation scientifique s'inscrit dans un continuum de la diffusion de la science, elle en est une des modalités. [...] J'établis que la diffusion des connaissances peut [...] s'opérer par une diffusion structurée où un petit nombre de lecteurs deviennent des ré-utilisateurs d'une information soutenue. La diffusion large, auprès d'un public indifférencié, par le moyen de rhétorique particulière n'est probablement qu'un cliché dénué de consistance. Dans les faits, c'est une large panoplie de pratiques de socio-diffusion de la science qu'il faudrait évoquer », Jacobi 1987, respectivement p. 8 et p. 163.

45. La contribution [Auroux 1994] a pour titre *Encyclopédies, bibliothèques et formalisation du savoir*.

46. Auroux 1994, p. 150.

47. *Ibid.*, p. 149.

48. Pour la période récente, Blanquet 1994 et Benoît 1992 proposent une synthèse des différentes positions sur les rapports entre information et connaissance en documentation.

proposé de comprendre l'information comme un « objet de discours » créé par l'utilisateur et l'indexation comme la pratique permettant de créer ces objets de discours. Quant à savoir si cet objet de discours permet à un utilisateur de s'informer ou de se former, nous ne disposons là d'aucun moyen de réponse.

En revanche, que l'on parle d'information ou de connaissance, la gestion des objets de discours en indexation pose des problèmes similaires. Il s'agit bien de maintenir une stabilité d'un bout à l'autre de la chaîne documentaire ; et c'est là que les remarques de Sylvain Auroux nous paraissent éclairantes.

En effet, qu'il s'agisse de représentations scientifiques ou de représentations « tout court », cet ensemble de représentations discursives que constituent les documents d'une bibliothèque ou d'un centre de documentation sont toutes singulières et ne correspondent pas *a priori* aux représentations, communes ou spécialisées, des utilisateurs, des lecteurs d'une bibliothèque ou d'un centre de documentation. En cela, la problématique de l'indexation rencontre celle de la vulgarisation : assurer une stabilité de la référence qui aille d'un auteur à un lecteur. Dès lors, l'indexation peut se penser sous l'angle de la vulgarisation scientifique sans que l'on ait forcément à se prononcer sur le caractère informatif ou pédagogique des accès qu'elle dispose. Son enjeu et sa finalité peuvent se laisser approcher sous l'angle du mode de diffusion des documents qu'elle réalise : c'est un mode de diffusion nécessairement élargi, nécessairement plus large que le contexte dans lequel les documents qu'elle propose ont été créés⁴⁹.

L'approche linguistique des discours de vulgarisation scientifique montre que le problème de la communication des savoirs vers des non-spécialistes n'est pas lié aux mots employés mais au contexte dans lequel ils sont utilisés et que tout l'effort du vulgarisateur porte précisément sur ce contexte. Autrement dit, l'approche linguistique de la vulgarisation fait émerger la dimension énonciative, complexe, de la communication vulgarisante et focalise l'attention sur les moyens linguistiques mis en œuvre pour opérer ce type de communication.

Il nous semble de même important de définir l'indexation en prenant en compte la diversité des foyers énonciatifs (relative aux auteurs comme aux indexeurs), diversité qui n'est pas réductible par l'adoption d'un système lexical réduit et unifié. Nous poserons donc, au cœur de la problématique du discours documentaire, cette pluralité d'univers de référence que doit permettre de traverser l'indexation. Dans les termes de Kripke, et comme nous le verrons ci-après, aussi bien en

⁴⁹. En ce sens, peut-être n'y a-t-il pas une grande différence entre la documentation, ou la bibliothèque, dite spécialisée et la documentation, ou la bibliothèque, dite encyclopédique : sans doute une nuance de degré peut-elle être établie (plus ou moins de domaines spécialisés en jeu, plus ou moins de types de lecteur différents).

documentation qu'en vulgarisation, il s'agit d'opérer le passage d'un monde possible, où la fixité de la référence est acquise, à une pluralité de mondes possibles, à travers lesquels la référence peut se trouver modifiée.

Face à cette problématique commune à l'indexation et à la vulgarisation scientifique, quels sont les moyens que peuvent mettre en œuvre l'une et l'autre des deux pratiques ?

Comme nous l'avons vu, dans le discours de vulgarisation scientifique, le lien est maintenu avec les sources scientifiques initiales par le biais des termes, qui cependant ne jouent plus le même rôle que celui qu'ils tiennent dans les discours spécialisés : ils fonctionnent comme des « termes pivots », dotés d'une double fonction référentielle. Ils réfèrent d'une part aux sources initiales dans lesquelles ils sont définis proprement comme termes. Ils réfèrent d'autre part aux discours dans lesquels ils apparaissent et fonctionnent alors comme indicateurs du début d'une explicitation scientifique qui va procéder par accumulation de procédés rhétoriques divers (schéma, exemple, métaphore, etc.)⁵⁰.

Nous posons comme hypothèse que l'indexation consiste, elle aussi, à repérer, dans un document, des types de mot qui tiennent moins un rôle sémantique qu'un rôle désignatif, et un rôle désignatif double : les descripteurs réfèreraient, dans ce cadre, à la fois aux autres discours et au discours d'où il est extrait, en pointant sur les segments textuels aptes à permettre la construction de l'information recherchée. La formulation de cette hypothèse rejoint celle que nous émettions au chapitre II où nous proposons de penser le descripteur à travers la notion de thème de discours telle que l'approche Marandin. À ce stade de la recherche et compte tenu des acquis issus de l'analyse linguistique du discours de vulgarisation, nous pouvons étoffer cette hypothèse : le descripteur, s'il doit fonctionner comme un thème de discours, doit, compte tenu de la diversité des univers de référence qu'il doit permettre de traverser, présenter des caractéristiques proches de celles du terme de la terminologie : c'est sur ces bases que nous conduirons notre étude du descripteur dans le chapitre V.

Le rapprochement, par analogie, entre indexation et vulgarisation scientifique dessine le rapport que peuvent entretenir en indexation « langage documentaire »

⁵⁰. Par exemple, l'extrait suivant s'articule autour du terme-pivot « diapir » : « La cartographie structurale nous a révélé, il y a une dizaine d'années, l'existence, sous le Moho, d'une montée de manteau plus ou moins tubulaire. Par analogie avec les diapirs de sel, qui sont des colonnes irrégulières de sel montant spontanément grâce à leur densité inférieure à celle des sédiments qu'ils traversent, ces cheminées de matériel mantélique furent nommées " diapirs " », *La Recherche*, n° 239, janvier 1992, p. 25, cité in Jacobi 1994.

(entendu comme ensemble de descripteurs extraits d'une collection de documents) et « discours documentaire » (entendu comme ensemble de textes organisés selon des principes que nous définirons) : le langage documentaire est au discours documentaire ce que la terminologie est au discours de vulgarisation, une chaîne de traces lexicales destinée à la construction de l'interprétation.

Bien évidemment, le rapprochement entre indexation et vulgarisation, s'il peut paraître éclairant, se heurte rapidement à des limites : la vulgarisation scientifique peut se concevoir comme un discours second lié *via* les termes à des discours premiers parce qu'elle produit elle-même des discours ; l'indexation ne produit pas de discours dans ce sens-là. Il reste à définir en propre le discours documentaire.

Le détour analogique par l'examen de l'approche linguistique de la pratique terminologique et de la pratique vulgarisatrice nous a semblé utile pour deux raisons :

- d'une part, il appuie la différence qui existe entre la perception de la pratique par elle-même et la perception de la pratique par des linguistes. La différence tient, dans les deux cas, à la focalisation extrême que les praticiens portent sur les mots au détriment des discours. En ce sens, il nous semble que les représentations que les indexeurs se font de leur pratique ne diffèrent pas des représentations d'autres praticiens ;
- d'autre part, il pointe ce qui, dans une pratique qui utilise des mots et des textes, reste souvent inaperçu : la dimension du discours sous ses deux aspects, textuel et énonciatif. Dans les deux cas, on voit que les termes entretiennent toujours des rapports étroits avec les textes dont ils sont issus. On voit aussi ce qui distingue mot et discours : ce ne sont pas les mots qui sont en terminologie « spécialisés » et en vulgarisation « didactiques » ; ce sont les discours dans lesquels ils sont employés qui sont spécialisés ou didactiques. De même, il importe de noter que ce ne sont pas les mots qui sont en indexation « informatifs » : c'est la façon dont ils se répondent de texte en texte qui peut créer de l'information, ou encore de la connaissance.

Après les avoir perçues uniquement par le biais d'autres pratiques, nous proposons ci-après de préciser les problématiques propres au discours documentaire en indexation.

I.4- Approche du discours documentaire

Avant de définir le « cahier des charges » du discours documentaire, nous tâcherons de préciser l'enjeu de cette notion pour une approche de l'indexation : en quoi l'hypothèse de la création, par l'indexation, d'un espace de discours particulier, intermédiaire entre l'espace des auteurs et l'espace des utilisateurs, est-elle nécessaire ?

I.4.1- Enjeu de la notion de discours documentaire

L'enjeu de la notion de discours documentaire, entendu comme espace d'organisation de documents, ne se pose que dans le cadre d'une certaine vision des institutions dans lesquelles l'indexation est pratiquée (bibliothèques et centres de documentation).

Sur ce point, on peut distinguer deux approches.

(i) La notion de discours documentaire est peu pertinente dans le cadre d'une approche qui considère la bibliothèque avant tout, et en vertu de son étymologie, comme un dépôt de livres, un lieu d'accumulation et de conservation, qu'il est nécessaire d'organiser simplement en raison du nombre des documents qui s'y trouvent. Les principes d'organisation sont donc dans ce cas établis *a posteriori* sur la base de critères qui peuvent être externes aux sources des documents ; ces principes sont le plus souvent « orientés utilisateurs ». Il nous semble que relèvent de cette approche tous les discours où l'enjeu de la documentation et par suite de l'indexation se situe dans le traitement massif, toujours plus rapide, de documents toujours plus nombreux. Dans ce cadre, tous les moyens sont bons pour traiter les documents, les « indexer » : méthodes informatiques, statistiques, linguistiques, cognitives, ensemble ou séparées, ne sont évaluées qu'à l'aune d'un seul critère, la faillite ou le succès d'une recherche documentaire menée par un utilisateur : nous retrouvons ici l'expression de l'hypothèse de service que nous avons dégagée au chapitre I.

(ii) À l'opposé, et toujours à gros traits, se situe une approche qui considère la bibliothèque comme auxiliaire de la pensée humaine au sens large et vague du terme. Cette approche est diversement illustrée. On la précisera en reprenant la vision de deux épistémologues des sciences, Auroux et Latour, qui mettent l'accent sur la nécessité d'organiser les savoirs en bibliothèque non plus *a posteriori*, en

rapprochant par des mots des documents déjà sélectionnés et constitués comme tels, mais *a priori*.

A- Auroux [1994] définit le rôle de la bibliothèque par rapport aux deux propriétés qui caractérisent, selon lui, le savoir scientifique : la formalisation et l'externalisation. C'est principalement sur la base de la seconde propriété d'externalité du savoir que la bibliothèque apparaît comme auxiliaire de la pensée : « le savoir a besoin pour fonctionner d'extension des capacités cognitives sous forme de bibliothèques, de discours de reformulation et de formalisation »⁵¹. Dans ce cadre, qui tient qu'il n'y a pas de savoir sans construction d'instruments externes, la bibliothèque constitue l'un de ces instruments. En ce sens, l'accès au savoir peut difficilement s'établir *a posteriori* sur la base d'une théorie, préexistante et externe aux sources, de la répartition des connaissances : « l'univers, soit réel soit intelligible, a une infinité de points de vue sous lesquels il peut être représenté et le nombre des systèmes possibles de la connaissance humaine est aussi grand que celui de ces points de vue »⁵².

On retrouve là une dialectique classique en documentation mais encore est-il nécessaire de la poser comme centrale, irréductible : si la bibliothèque est un instrument consubstantiel au savoir, son enjeu ne peut être celui de trouver un mode de représentation universelle d'un savoir, ce savoir n'existant pas sans, entre autres, la bibliothèque elle-même. Il s'agit plutôt de trouver des modes de représentation qui ménagent la diversité des points de vue qui, eux, permettront l'émergence du savoir.

Ce texte d'Auroux nous semble mettre en valeur la nécessité de penser la bibliothèque en termes d'espace spécifique et particulièrement en termes d'espace de représentations spécifique.

B- On retrouve chez Latour [1996], dans le cadre d'une problématique différente, la même expression de cette nécessité :

On comprend alors que les institutions comme les bibliothèques, les laboratoires, les collections ne soient pas de simples moyens dont on pourrait se dispenser aisément, sous prétexte que les phénomènes parleraient par eux-mêmes à la seule lumière de la raison. Additionnés les uns aux autres, ils composent les phénomènes qui n'ont d'existence que par cet étallement à travers la série des transformations.
Latour, 1996, p. 41.

L'argumentation que Latour défend contre la vision de la bibliothèque comme forteresse de papier ou comme empire de signes se déploie autour de la notion de

⁵¹. Auroux 1994, p. 147.

⁵². *Ibid.* p. 145.

« centre de calcul »⁵³.

En qualifiant la bibliothèque de centre de calcul, Latour cherche à montrer l'articulation qu'opère la bibliothèque avec le « monde », avec son extérieur au sens large : elle est l'un des éléments qui participent à la construction de ce monde, à la construction de la réalité⁵⁴. Là aussi elle se présente comme un auxiliaire, précieux, indispensable, de la pensée humaine. Son caractère indispensable tient à ce que, en tant que centre de calcul, la bibliothèque s'approche comme un lieu de transformation.

La transformation s'y réalise par un double mouvement de réduction et d'amplification⁵⁵ :

- la réduction est liée au fait que la bibliothèque ne manipule que des « inscriptions » dans les termes de Latour : par exemple, photo d'une situation, compte rendu d'une expérience scientifique et non situation elle-même ou expérience scientifique elle-même ; il s'agit, dans ces derniers cas, de « phénomènes » qui évoluent dans des lieux périphériques au lieu central du calcul ;
- l'amplification est liée au fait que la bibliothèque manipule des objets de même nature, susceptibles d'échanger entre eux des propriétés. Appréhendés sous une même forme, celle des inscriptions, les « phénomènes » au sens de Latour deviennent commensurables. Là où les phénomènes ne se donnent qu'en ordre dispersé, la bibliothèque donne la possibilité de les appréhender sous une forme unifiée : là encore il est question de regard, de « cohérence optique », dit Latour⁵⁶.

Dans l'approche de Latour, l'information se définit comme une relation entre les lieux périphériques (lieux des phénomènes) et le centre de calcul (lieu des inscriptions) ; en ce sens, l'information n'est pas un signe isolé, mais un signe chargé de la matière des phénomènes⁵⁷. Elle est une inscription réduite par rapport à son lieu d'origine et amplifiée dans le nouveau lieu où elle se trouve : c'est par cette

⁵³. La notion fait l'objet d'une description précise dans le cadre du modèle de Latour [1989] : on ne peut ici entrer dans les détails. Nous reprenons simplement l'approche qu'en propose Latour dans le texte qui nous occupe : « Pour comprendre un centre de calcul, il faut donc tenir du doigt l'ensemble du réseau des transformations qui relie chaque inscription au monde et qui relie ensuite chaque inscription à toutes celles qui lui sont devenues commensurables par la gravure, le dessin, le récit, le calcul ou, plus récemment, la numérisation », Latour 1996, p. 36.

⁵⁴. *Ibid.*, p. 28.

⁵⁵. *Ibid.*, p. 25-27.

⁵⁶. *Ibid.*, p. 33 : « Des informations éparses, provenant d'instruments épars peuvent s'unifier en une seule vision parce que les inscriptions possèdent toutes la même cohérence optique ».

⁵⁷. *Ibid.*, p. 25.

transformation que s'établit, selon Latour, la connaissance.

Ce qui permet l'émergence de la connaissance, c'est la « cohérence optique » à l'œuvre au niveau du centre de calcul. Cette cohérence optique n'est pas établie sur la base des signes proprement dits mais sur la base d'une compatibilité entre les inscriptions, qui permet aux inscriptions d'échanger entre elles des propriétés⁵⁸, de « capitaliser », dit Latour. En ce sens, l'inscription, ou l'information (les termes ne sont pas toujours distingués dans ce texte de Latour), se situe entre les mots et les choses ou plutôt traverse les deux mondes des mots et des choses :

Ces inscriptions circulent dans les deux sens, seul moyen d'assurer la fidélité, la fiabilité, la vérité entre le représenté et le représentant. Comme elles doivent à la fois permettre la mobilité des rapports et l'immutabilité de ce qu'elles transportent, je les appelle des « mobiles immuables » afin de bien les distinguer des signes. En effet, lorsqu'on les suit, on se met à traverser la *distinction usuelle entre mots et choses*, on ne voyage plus dans le monde, mais aussi dans les matières différentes de l'expression.

Latour, 1996, p. 38.

L'image de la bibliothèque que construit Latour est celle d'un laboratoire de transformation. Il compare d'ailleurs la bibliothèque à un laboratoire du CERN où sont « redistribuées les propriétés des phénomènes qui n'existent nulle part ailleurs et que savent saisir, repérer, amplifier des détecteurs géants construits pour l'occasion »⁵⁹.

De ce texte foisonnant d'images, dans tous les sens du terme, nous retiendrons que la bibliothèque peut s'appréhender en termes d'espace de création de la réalité, ou plutôt d'une réalité. Elle est le lieu d'une rencontre et d'un échange entre phénomènes et inscriptions, entre les mots et les choses, un lieu où, par le format unique des inscriptions, la diversité des phénomènes a quelque chance de se trouver unifiée, pour peu qu'une « cohérence optique », qu'un « regard » puisse les faire voir comme semblables.

Dans ce cadre, l'indexation se conçoit comme ce « regard » qui permet aux phénomènes rapportés par les inscriptions de trouver une cohérence optique. Sur ce point, Latour souligne la nécessité de maintenir une distinction entre monde des signes, monde des inscriptions et monde des phénomènes, puisque seule cette distinction peut permettre de penser leurs interrelations⁶⁰. Nous retrouvons en outre chez Latour une autre formulation de notre hypothèse concernant l'approche de l'information en termes d'objet de discours : sa notion de « mobile immuable »

⁵⁸. *Ibid.*, p. 38.

⁵⁹. *Ibid.*, p. 39.

⁶⁰. *Ibid.*, p. 41 : « La vérédiction ne vient pas de la superposition d'un énoncé et d'un état du monde, mais provient plutôt du maintien continu des réseaux, des centres et des mobiles immuables qui y circulent ».

souligne la nécessité de penser l'information sous la double face d'une permanence et d'une instabilité référentielle.

Si nous avons largement mobilisé, depuis le début de ce chapitre, des auteurs issus de diverses disciplines, qui ne présentent pas toujours un lien direct avec notre sujet, si nous avons tenté de faire croiser ces différents regards, en prenant les risques de la démarche analogique, c'est que la dimension discursive de l'indexation, l'organisation spécifique des discours qu'elle réalise, n'est guère présente dans les approches classiques de l'indexation : elle reste entièrement à définir.

Pour contribuer à la définition de cet espace de discours documentaire, nous devons d'abord cerner ses différentes facettes et tenter de dégager son enjeu. Nous nous proposons désormais d'établir des référentiels (des cadres d'analyse, modèles ou théories) qui permettent de travailler ces différentes problématiques : nous les synthétisons au sein d'un « cahier des charges » du discours documentaire qui nous permettra de cerner le type de cadres théoriques nécessaires à retenir pour poursuivre notre recherche.

I.4.2- Le discours documentaire : « cahier des charges »

Nous avons pris comme point de départ la notion de langage documentaire, non pas tel ou tel langage documentaire (classification, thésaurus, liste de vedettes-matières, etc.), mais plutôt son principe : la représentation d'un texte établie au travers de mots considérés en eux-mêmes.

C'est d'abord dans la notion même de langage documentaire que la dimension discursive de l'indexation nous est apparue. En effet, la notion de langage documentaire, confuse en ce qu'elle ne distingue pas ses éléments, rend opaque les deux dimensions qui la constituent comme discours : la dimension textuelle et la dimension énonciative. Pourtant, la dimension textuelle est présente au tout début de l'indexation : les descripteurs qu'elle utilise sont toujours et d'abord des mots issus de discours, des mots du discours. De même, la dimension énonciative de l'indexation est-elle inscrite dans l'espace documentaire proposé à l'utilisateur. C'est en considérant les indexeurs comme formant un cercle étroit de locuteurs, dans lequel les mots peuvent trouver une certaine stabilité désignative, qu'apparaît l'enjeu de l'élargissement de ce cercle en indexation ; sur ce point, la problématique de l'indexation rejoint celle de la vulgarisation.

Penser l'indexation par le seul biais du langage documentaire paraît alors fortement réducteur : on n'y voit ni la dimension des textes ni celle des différents acteurs. Il nous semble qu'il faut alors observer l'amont du langage documentaire comme l'aval :

- en amont : c'est la sélection des sources elle-même qui doit retenir l'attention, car les descripteurs sont des mots issus de ces sources-là, et non de sources antérieures, passées, oubliées, jugées, le temps de constituer un langage documentaire, représentatives de toutes les sources à venir ;
- en aval : c'est l'organisation des documents, la façon de les soumettre au « regard », pour les rendre commensurables, pour permettre entre eux une circulation, qui mérite d'être observée ; là encore les mots ne sont pas seuls en jeu.

Sélection des sources et organisation des documents engagent une stratégie qui porte sur les discours eux-mêmes. Cette stratégie est, selon nous, double : il y a une stratégie qui concerne les sources (quels textes choisir ?) et une stratégie qui concerne les documents (comment montrer les documents ?).

Il y a plusieurs cadres d'analyse pour travailler ces deux aspects de la stratégie documentaire. Nous en avons retenu deux, sur la base de leur compatibilité avec une approche linguistique du lexique et de la référence, les deux problématiques qui nous paraissent centrales en indexation. C'est ainsi que pour donner quelques fondements à l'indexation comme stratégie d'exploration de sources, nous nous référerons au modèle des formations discursives de Foucault (II) ; pour donner quelques fondements à l'indexation comme stratégie d'exposition des documents, nous emprunterons des éléments à la théorie des mondes possibles établie par Kripke (III).

II- Stratégie d'exploration des sources en indexation

Si l'indexation peut se penser sous l'angle d'un discours, entendu comme « ensemble d'énoncés et/ou de textes, possédant une organisation thématique, normative, structurale »⁶¹, la question de l'indexation comme discours pose d'abord celle du type d'organisation qu'elle établit à partir de textes ou d'énoncés qui ne sont pas donnés *a priori* comme semblables, mais qu'elle donne à voir comme un « ensemble » (la collection documentaire).

Le discours documentaire est alors la forme du regroupement que réalise l'indexation à partir de sources hétérogènes. Cette forme de regroupement trouve généralement deux modes d'inscription : d'une part, sur un plan textuel, dans le document (regroupement, éclatement ou maintien des sources originelles⁶²), d'autre part, sur un plan lexical, dans les descripteurs (qui matérialisent alors les liens entre documents).

Insistons sur le fait que ces modes d'inscription ne sont que des conséquences de l'opération de transformation discursive réalisée par l'indexation en amont, les conséquences du passage d'un type de discours (propre à une source, produite dans un certain contexte, promise à un certain usage) à un autre type de discours (propre au document, inséré dans un autre contexte, destiné à d'autres usages). Pour reprendre un exemple déjà cité, l'indexation est appréhendée ici comme ce qui fait passer un texte de la catégorie discursive « brevet » à la catégorie discursive « veille informative » quand ce texte est donné à lire comme porteur d'informations sur des innovations, et non plus comme attestation juridique protégeant un inventeur.

Il y a, schématiquement, deux façons de sélectionner et de regrouper des textes : soit *a posteriori* (c'est par exemple la notion classique de regroupement thématique), soit *a priori* (c'est par exemple la notion de formation discursive chez Foucault). Nous avons dégagé précédemment la nécessité, en indexation, de disposer d'un mode d'organisation qui soit établi *a priori*, sur d'autres critères que celui du « contenu » lui-même d'une source : paradoxalement, c'est là le seul moyen de ménager des points de vue différents sur des documents, tout en ayant la possibilité de déterminer des contours.

C'est en ce sens que la notion de « formation discursive » proposée par Foucault nous paraît nécessaire à disposer au centre de cette recherche comme « horizon

⁶¹. Souchard 1995, p. 258.

⁶². Voir précédemment le compte rendu d'expérience, chapitre III § I.3., ou encore l'annexe 3 pour le détail.

théorique ».

La notion de formation discursive pourra paraître bien trop puissante pour couvrir les aspects très triviaux des problèmes documentaires. Il faudrait que les documentalistes et les bibliothécaires se livrent à un véritable travail scientifique sur les collections qui, s'il s'est fait pendant longtemps et continue à se pratiquer au sein de certaines bibliothèques⁶³, n'est pas toujours envisageable dans toutes les structures. Reste que le problème de la sélection et du regroupement, le problème plus général de la constitution de corpus s'avère, comme nous le montrerons, si déterminant pour l'attribution ou l'extraction *in fine* des descripteurs, qu'il semble nécessaire de poser au moins un cadre théorique à des modes de résolution qu'il restera à définir.

C'est dans cette optique que nous présentons dans ce paragraphe :

- d'une part, la notion de « système-archive » établie par Foucault, élément de son modèle des formations discursives qui nous paraît particulièrement adapté à une approche du discours documentaire à la recherche de fondements théoriques ;
- d'autre part, les modes d'exploration des sources habituellement utilisés par les professionnels et les problèmes qu'ils posent au regard de la notion de système-archive.

Le recours au modèle de Foucault pour examiner la notion de discours documentaire nous a été inspiré, d'une part, par les travaux réalisés en analyse de discours⁶⁴, d'autre part, par ceux réalisés en indexation par Suzanne Bertrand-Gastaldi⁶⁵. Le modèle de Foucault a en outre été retenu dans la mesure où il nous paraissait compatible avec l'approche linguistique du lexique et de la référence qui nous sert de « toile de fond » théorique ; en effet, on lit par exemple dans Foucault [1966] :

Je voudrais montrer que le discours n'est pas une mince surface de contact, ou d'affrontement entre une réalité et une langue, l'intrication d'un lexique et d'une expérience ; je voudrais montrer sur des exemples précis, qu'en analysant les discours eux-mêmes, on voit se desserrer l'étreinte apparemment si forte des mots et des choses, et se dégager un ensemble de règles propres à la pratique discursive. Ces règles définissent non point l'existence muette d'une réalité, non point l'usage canonique d'un vocabulaire, mais le régime des objets.
Foucault 1966, p. 66.

C'est dans ce cadre que Foucault se propose de conduire une histoire des objets discursifs.

⁶³. Notamment à la Bibliothèque nationale de France, qui développe un axe important de recherche sur les collections spécialisées ; un tel travail scientifique se mène aussi dans la plupart des bibliothèques dites d'étude.

⁶⁴. Marandin 1984 par exemple, mais aussi Jacobi 1988 par exemple.

⁶⁵. Bertrand-Gastaldi 1989 et 1993, par exemple.

II.1- La notion de « système-archive » dans le modèle de Foucault

Nous présenterons la notion de « système-archive » après avoir rappelé le projet de Foucault sous les aspects qui nous paraissent pertinents pour l'approche de cette notion dans notre recherche. Nous dégagerons, en fin de paragraphe, l'enjeu de la notion de « système-archive » en indexation.

II.1.1- Projet de Foucault

Loin d'être spécialiste de Foucault⁶⁶, nous ne reprendrons sommairement que quelques éléments qui, de la problématique ou de la méthode, nous paraissent importants pour situer l'utilisation que nous ferons des concepts qu'il a établis.

A- Problématique

L'une des problématiques de Foucault peut se trouver résumée dans la question suivante : pourquoi ne peut-on pas dire que Darwin parle de la même chose que Diderot⁶⁷ ? Ou, sous une forme positive, au nom de quoi relie-t-on des auteurs qui ne se connaissent pas « dans une trame dont ils ne sont pas maîtres ?⁶⁸ ». Foucault, qui pose la question de la légitimité d'un corpus, s'attache alors à montrer l'insuffisance de la notion classique de « thème »⁶⁹. Qu'il s'agisse de son travail sur la folie, le pouvoir ou la répression, le principe reste toujours de ne pas considérer d'emblée ces « mots » comme constituant des invariants, sur la base desquels des textes pourraient être rapprochés. Il s'agit non de dégager des permanences mais d'isoler les « lois » de coexistence des textes maintenus dans leur singularité.

La base de travail reste fournie par les regroupements thématiques habituellement effectués, mais l'enjeu consiste à s'en défaire, à déconstruire ces regroupements :

Certes, je prendrai pour repère initial des unités toutes données (comme la psychopathologie, ou la médecine, ou la philosophie politique) ; mais je ne me placerai pas à l'intérieur de ces unités douteuses pour en étudier la configuration interne ou les secrètes contradictions. Je ne m'appuierai sur elles que le temps de me demander quelles unités elles forment ; de quel droit elles peuvent revendiquer un domaine qui les spécifie dans l'espace et une continuité qui les individualise dans le temps ; selon quelles lois elles se forment ; sur fond de quels événements elles se découpent ; et si finalement elles ne sont pas, dans leur individualité acceptée et

⁶⁶. Nous nous appuyerons dans ce paragraphe essentiellement sur Foucault 1969 et sur les lectures qu'en ont faites Marietti 1985 [1974] et Marandin 1984, 1993, 1997.

⁶⁷. Foucault 1969, p. 166.

⁶⁸. *Ibid.*, p. 167.

⁶⁹. « On aurait tort de chercher dans l'existence des thèmes les principes d'individuation d'un discours ». *Ibid.*, p. 51.

quasi institutionnelles, l'effet de surfaces d'unités plus consistantes.
Foucault, 1969, p. 142.

Le matériau de base est donc constitué des « choses dites », dont il s'agit de dégager les règles de formation.

La notion de règles de formation des discours s'articule sur celle d'*a priori historique* : « l'histoire des choses effectivement dites jouit d'un *a priori* (historique) auquel l'archéologie assigne le rôle de rendre compte des énoncés dispersés »⁷⁰. L'*a priori historique* qui donne la possibilité d'identifier des règles de formations discursives est un *a priori* qui ne concerne ni le sens ni la référence d'un énoncé mais plutôt sa « condition de réalité » (sa « positivité », dit encore Foucault) ; cet *a priori* est historique car il se donne dans le cadre d'une histoire spécifique.

La notion d'*a priori historique* peut aussi être comprise, et c'est ce sens que privilégie Marandin⁷¹, comme un « préconstruit », ce préconstruit pouvant valoir d'hypothèse pour la constitution d'un corpus⁷².

Il ne relève pas d'une théorie de la langue mais d'« une théorie du fonctionnement de la langue dans une formation sociale »⁷³. En cela, la notion de préconstruit se distingue de celle du « déjà-dit » qui porte, elle, sur le contenu. C'est ainsi que, dans le modèle de l'analyse de discours proposé par Pêcheux, l'interdiscours se comprend aussi en termes de préconstruit⁷⁴.

Reste que, dans les corpus établis sur la base d'un *a priori historique* ou encore sur la base d'un préconstruit, des effets d'invariance, comme des effets de déjà-dits, peuvent se rencontrer ; mais il s'agit là uniquement d'effets qu'on ne saurait prendre pour des principes de regroupement :

Cette forme de positivité [*a priori historique*] (et les conditions d'exercice de la fonction énonciative) définit un champ où peuvent éventuellement se déployer des identités formelles, des continuités thématiques, des translations de concepts, des jeux polémiques.

Foucault, 1969, p. 167.

B- Méthode

Pour mener son projet d'identification des règles de formations discursives, Foucault se dote de trois méthodes d'analyse⁷⁵ ; seule l'une d'entre elle sera ici

⁷⁰. Marietti 1985 [1974] p. 155.

⁷¹. Marandin 1984, 1993, 1997.

⁷². Marandin 1984 oppose un jugement de ressemblance entre textes qui se ferait au niveau des « formes des énoncés » à un jugement de ressemblance qui s'élaborerait au niveau d'un préconstruit (appelé alors « pré-discursif »)

⁷³. Marandin 1993, [p. 10].

⁷⁴. Marandin 1997, p. 12.

⁷⁵. Il s'agit de mener conjointement une analyse des pratiques discursives, une analyse des

évoquée.

La notion de « pratique discursive » constitue une méthode d'analyse que Foucault a établie pour « suivre la formation des savoirs »⁷⁶. Les discours sont, dans ce cadre, appréhendés comme des pratiques discursives, obéissant à ce titre à des lois de formation. Ces lois sont dégagées sur la base d'un examen des stratégies propres à une pratique. Par stratégie, Foucault entend aussi bien les « thèmes » spécifiques à une pratique que les « théories » élaborées au sein d'une pratique⁷⁷. Thèmes et théories ne sont pas considérés comme points de départ mais comme points d'arrivée de l'analyse : l'analyse des pratiques discursives doit faire émerger un « niveau préconceptuel » par rapport aux discours eux-mêmes. La méthode d'analyse vise donc bien à capter le régime d'existence des objets de discours⁷⁸.

Thèmes et théories propres à une pratique sont alors extraits de leur domaine de validité courant, ils sont décontextualisés et reconsidérés dans un cadre nouveau :

Ce qui a été modifié [par l'enquête archéologique], c'est le rapport de ces affirmations [stratégies au sens de Foucault] à d'autres propositions, ce sont leurs conditions d'utilisation et de réinvestissement, c'est le champs d'expérience, de vérifications possibles, de problèmes à résoudre auquel on peut les référer.
Foucault, 1969, p. 136.

Ainsi, par exemple, l'objet « folie » se constitue-t-il au sein de pratiques diverses touchant l'hospitalisation, l'internement, l'exclusion sociale, la jurisprudence, les normes du travail industriel et de la morale, etc.⁷⁹. Il se constitue au gré des transformations que lui impriment ces pratiques diverses :

D'une façon générale, définir un ensemble d'énoncés dans ce qu'il a d'individuel consisterait à décrire la dispersion de ces objets ; saisir tous les interstices qui les séparent, mesurer les distances qui règnent entre eux - en d'autres termes, formuler leur loi de répartition.
Foucault, 1969, p. 46-47.

La méthode d'analyse des pratiques discursives procède donc en deux phases : une phase de reconstitution d'une épistémologie (« la somme de ce que l'on a cru vrai ») et une phase de reconstitution du savoir (« ce dont on a pu effectivement parler »). Dans ce cadre, le savoir ne réside pas seulement dans les propositions

relations de pouvoir et une analyse des modes de reconnaissance des sujets.

⁷⁶. Foucault 1984, p. 12 : « Un déplacement théorique m'avait paru nécessaire pour analyser ce qui était souvent désigné comme le progrès des connaissances : il m'avait conduit à m'interroger sur les formes de pratiques discursives qui articulaient le savoir ». C'est nous qui soulignons.

⁷⁷. Marietti 1985 [1974] p. 41.

⁷⁸. *Ibid.*, p. 37.

⁷⁹. Foucault 1969, p. 234 : « Cette pratique discursive, elle était investie dans la médecine certes, mais tout autant dans les règlements administratifs, dans des textes littéraires ou philosophiques, dans la casuistique, dans les théories ou les projets de travail obligatoire ou d'assistance aux pauvres ».

scientifiques : il y a élargissement nécessaire à d'autres champs de discours⁸⁰. S'il s'agit de s'attacher à ce qui fonde les thèmes ou les théories propres à une pratique, l'enquête archéologique ne consiste pas pour autant à restituer l'acte de fondation de ces thèmes ou théories⁸¹.

L'analyse des discours en tant que pratiques discursives se situe donc à un niveau particulier, qui n'est ni le stade de l'avant-verbalisation d'une théorie ni celui de la verbalisation elle-même. Ce niveau que l'analyse reconstitue est celui qui permet de capter le moment où l'objet de discours apparaît (sous la forme d'un événement) et se fixe (sous la forme d'une chose) : c'est ce niveau propre à une pratique spécifique que Foucault nomme un « système-archive »⁸².

II.1.2- Notion foucauldienne de formations discursives

La notion de « formation discursive », intimement liée à celle d'« archéologie du savoir » et de « positivité des discours », se laisse aborder de multiples façons⁸³. Dans la présentation qui suit, nous avons privilégié, en les simplifiant, trois approches : une approche en termes d'« espace », une approche en termes d'« archive » et une approche en termes de « domaine ». Sous ces trois angles se dégagent des moyens pour penser la notion de corpus, et, plus précisément, la problématique du regroupement des textes en indexation.

A- Formation discursive et création d'espace discursif

Une formation discursive permet de considérer un ensemble de discours comme formant un « espace limité de communication »⁸⁴.

80. Marietti 1985 [1974], p. 123 : « Il faut comprendre que le savoir ne réside pas seulement dans les propositions scientifiques : l'enquête archéologique trouve son territoire aussi bien dans les textes littéraires ou philosophiques, dans les fictions, les réflexions, les récits, les règlements institutionnels, les décisions politiques ».

81. Foucault 1969, p. 235 : « Les formations discursives, ce ne sont donc pas les sciences futures dans le moment où [elles sont] inconscientes d'elles-mêmes ».

82. « Ce qui a pu être dit obéit à une loi que représente l'archive qui est le système qui régit l'apparition des énoncés comme événements singuliers. Définissant le système de l'énonciabilité de l'énoncé-événement, définissant le mode d'actualité de l'énoncé-chose, l'archive n'est autre que le système de son fonctionnement. Système général de la formation et la transformation des énoncés, l'archive fait surgir le niveau d'une pratique des énoncés dans leur valeur d'événements et de choses », Marietti 1985 [1974], p. 155.

83. Ce n'est pas la moindre des difficultés que présente le texte de Foucault ; sur ce point, Marandin [1979, p. 48] relève : « Le texte de Foucault est diffus ; les notions sont définies par des suites de négations. Ce qui en rend la lecture difficile, et problématique toute tentative de description effective (sur des discours différents que ceux que Foucault a choisis et par quelqu'un qui n'est pas Foucault) ».

84. Foucault 1969, p. 166.

Considérée en termes d'espace discursif, une formation discursive présente les caractéristiques suivantes :

- c'est un espace polyphonique :
devant « rendre compte des énoncés dans leur dispersion, dans toutes les failles ouvertes par leur non-cohérence, dans leur chevauchement et leur remplacement réciproque, dans leur simultanéité qui n'est pas unifiable, dans leur succession qui n'est pas déductible », l'espace discursif est, dans l'enquête archéologique, composé de sources textuelles d'origine nécessairement hétérogène : textes juridiques, littéraires, réglementaires, scientifiques, etc. De la même façon que le regroupement par thèmes avait paru douteux, le regroupement par genre ou par auteur est ici écarté⁸⁵ ;
- c'est un espace réglé :
une formation discursive est déterminée par une même « loi de répartition », c'est-à-dire par un ensemble de règles qui décrivent, pour une période donnée, et dans une aire géographique donnée, les conditions d'existence des énoncés, c'est-à-dire leurs conditions de production et d'interprétation. Ces règles sont de nature exclusivement discursive : « s'il y a des choses dites - et celles-là seulement -, il ne faut pas en demander la raison immédiate aux choses qui s'y trouvent dites ou aux hommes qui les ont dites, mais au système de la discursivité, aux possibilités et aux impossibilités énonciatives qu'il ménage » ;
- c'est un espace historiquement réglé :
la notion de formation discursive montre que le discours n'a pas seulement un sens et une vérité, il a aussi une histoire et une « histoire spécifique », propre à une « discipline »⁸⁶, ou mieux à un « domaine » (voir ci-après). En ce sens, le principe de la formation discursive relève d'un *a priori historique* qui est à distinguer des *a priori* formels : le premier ne peut rendre compte des seconds mais « permet de comprendre comment les *a priori* formels peuvent avoir dans l'histoire des points d'accrochage, des lieux d'insertion, [...] et de comprendre comment cette histoire peut

⁸⁵. Sur l'absence de prise en compte de la notion d'auteur et d'œuvre chez Foucault, voir Marietti 1985 [1974], p. 99-120.

⁸⁶. Cependant, il n'y a pas de relation bi-univoque entre disciplines instituées et formations discursives, ne serait-ce que parce que l'archéologie du savoir se tient à distance de l'épistémologie, voir sur ce point Foucault 1969, p. 232-239.

être non point contingence [...] mais régularité spécifique ».

L'approche des formations discursives que propose Foucault à travers la notion d'espace limité de communication montre que l'exploration parmi les sources textuelles peut se faire sur d'autres bases que sur les notions de thème, genres textuels et auteurs. Il faut alors étudier les conditions de production historiques des écrits. L'approche en termes d'archive souligne elle qu'il s'agit là d'une **nécessité**.

B- Formation discursive et création d'un système-archive

La notion d'archive est définie ainsi par Foucault :

Au lieu de voir s'aligner, sur le grand livre mythique de l'histoire, des mots qui traduisent en caractères visibles des pensées constituées avant et ailleurs, on a, dans l'épaisseur des pratiques discursives, des systèmes qui instaurent les énoncés comme des **événements** (ayant leurs conditions et leur domaine d'apparition) et des **choses** (comportant leur possibilité et leur champ d'utilisation). Ce sont tous ces systèmes d'énoncés (événements pour une part, et choses pour une autre) que je propose d'appeler *archive*.

Foucault, 1969, p. 169, (c'est nous qui soulignons).

On retiendra notamment de cet extrait que le « système-archive » gère les énoncés sous leur double face d'« événement » et de « chose ». On pourra voir dans le type « événement » le contexte d'énonciation, c'est-à-dire les traces discursives que garde l'énoncé décontextualisé, et dans le type « chose », la possibilité pour un énoncé d'être autonome par rapport à son contexte de production, et prêt en cela à être « détourné » de son usage initial. Mais l'archive telle que Foucault la définit présente encore, dans le cadre de notre recherche, cet intérêt qu'elle correspond à un niveau particulier entre la langue et le corpus, un niveau « qui définit le mode d'actualité de l'énoncé-chose » :

Entre la *langue* qui définit le système de construction des phrases possibles, et le *corpus* qui recueille passivement les paroles prononcées, l'*archive* définit un niveau particulier : celui d'une pratique qui fait surgir une multitude d'énoncés comme autant d'événements réguliers, comme autant de choses offertes au traitement et à la manipulation.

Foucault, 1969, p. 171.

Fondé sur des règles permettant « aux énoncés à la fois de subsister et de se modifier régulièrement », le système-archive se donne « comme un système général de la formation et de la transformation des énoncés ». Reste que l'archive d'une époque ne se capte jamais dans sa globalité : « elle se donne par fragments, régions et niveaux »⁸⁷.

Par la notion de système-archive apparaît la possibilité de créer un niveau

⁸⁷. *Ibid.*, p. 171.

intermédiaire de discours où un texte apparaît à la fois en tant que « source » et comme « document », pour reprendre les termes que nous avons précédemment proposés. Il nous semble que l'indexation devrait s'attacher à créer un tel niveau qui donne à voir **la formation** des énoncés dans le cadre qui les constitue mais aussi **leur transformation**, quand ils sont rapprochés d'autres énoncés au sein de formations discursives.

Si l'indexation parvient à être la « transformation réglée de ce qui a été déjà écrit »⁸⁸, alors elle constitue des « domaines de savoir ».

C- Formation discursive et création d'un domaine de savoir

La notion de « savoir » est définie par Foucault⁸⁹ sous les quatre dimensions suivantes :

- une dimension « sémantique » : un savoir, c'est « ce dont on peut parler dans une pratique discursive » ; c'est l'ensemble des objets discursifs manipulés par une pratique. Certains de ces objets pourront acquérir un statut scientifique mais un domaine de savoir ne saurait être réduit à ceux-là. Ainsi, pour reprendre l'exemple de Foucault, « le savoir de la psychiatrie, au XIX^e siècle ce n'est pas la somme de ce qu'on a cru vrai, c'est l'ensemble des conduites, des singularités, des déviations dont on peut parler dans le discours psychiatrique »⁹⁰ ;
- une dimension énonciative, celle des locuteurs : un savoir, c'est un « espace dans lequel un sujet peut prendre position pour parler des objets auxquels il a affaire dans son discours » ; en ce sens, la légitimité de la prise de parole dans un discours est élargie au-delà des seules autorités reconnues compétentes. Le sujet de discours n'est plus appréhendé comme un « auteur » propriétaire d'un savoir mais comme un « sujet situé et dépendant » ;
- une dimension intertextuelle, celle des relations entre textes : un savoir, c'est un « champ de coordination et de subordination des énoncés où les concepts apparaissent, se définissent, s'appliquent et se transforment ». En ce sens, et toujours en reprenant un exemple donné par Foucault, « le savoir de l'histoire naturelle, au XVIII^e siècle ce n'est pas la somme de ce qui a été dit, c'est l'ensemble des modes et des emplacements selon lesquels on peut intégrer au déjà

⁸⁸. *Ibid.*, p. 183.

⁸⁹. *Ibid.*, p. 238.

⁹⁰. *Id.*

dit tout énoncé nouveau » ;

- une dimension fonctionnelle, celle de l'usage : un savoir « se définit par des possibilités d'utilisation et d'appropriation offertes par le discours ». Ainsi, selon Foucault, le savoir de l'économie politique à l'époque classique « ce n'est pas la thèse des différentes thèses soutenues mais c'est l'ensemble de ses points d'articulation sur d'autres discours ou sur d'autres pratiques qui ne sont pas discursives »⁹¹.

En approchant la notion de formation discursive par le biais de la notion de domaine de savoir, Foucault souligne que le savoir ne peut se concevoir que dans le cadre de pratiques discursives données : « il n'y a pas de savoir sans une pratique discursive définie ; et toute pratique discursive peut se définir par le savoir qu'elle forme »⁹². De même il n'y a pas, non plus, d'objet de discours sans domaine de savoir, sans inscription dans une pratique discursive spécifiée.

En ce sens, si l'indexation, pour « fournir des informations », doit permettre de créer des objets de discours, ces objets de discours doivent nécessairement se laisser construire dans des domaines de savoir régis par des pratiques discursives spécifiques. En ce sens, un langage documentaire ne peut impunément remplacer un terme par un autre, sans que tout le pan de son inscription dans un discours ne disparaisse. Par l'approche du discours documentaire dans le cadre du modèle de Foucault, l'enjeu de l'indexation n'est définitivement plus du côté des mots.

Nous avons retenu de Foucault [1969] trois approches de la notion de formation des discours (approches en termes d'espace, d'archive et de domaine de savoir) qui devraient nous permettre de problématiser la notion de discours documentaire telle que nous la postulons.

II.1.3- Enjeu du modèle en indexation

Foucault prend le soin de préciser que la notion d'archive qu'il établit n'a rien à voir avec la « somme de tous les textes » que pourrait conserver une bibliothèque : l'archive « ne constitue pas la bibliothèque sans temps ni lieu de toutes les bibliothèques »⁹³.

Il y a, dans ces pages, une critique implicite de la conception classique des textes en

⁹¹. *Id.*

⁹². *Ibid.*, p. 238-239.

⁹³. *Ibid.*, p. 171.

bibliothèque, critique d'une conception naïve car purement réaliste qu'il nous a précédemment semblé important de dégager pour pouvoir s'en dégager. La bibliothèque ne considère le plus souvent les textes que comme les reflets des choses, des événements, sans pleinement prendre en compte l'épaisseur discursive des objets qu'elle détient :

[L'archive] c'est plutôt, c'est au contraire ce qui fait que tant de choses dites, par tant d'hommes depuis tant de millénaires, n'ont pas surgi par les seules lois de la pensée, ou d'après le seul jeu des circonstances, qu'elles ne sont pas simplement la signalisation, au niveau des performances verbales, de ce qui a pu se dérouler dans l'ordre de l'esprit ou dans l'ordre des choses ; mais qu'elles sont apparues grâce à tout un jeu de relations qui caractérisent en propre le niveau discursif.
Foucault, 1969, p. 170.

Foucault attire notre attention sur la non-évidence à la fois de l'existence des textes et de leur coexistence. Que faire de cette mise en garde d'une part et de l' ancestrale ignorance de la spécificité discursive des objets en bibliothèque d'autre part ? D'un côté, il nous paraît délicat d'ignorer le rôle des pratiques discursives tel que Foucault l'a mis en valeur ; d'un autre côté, il paraît tout simplement impossible de conduire, en bibliothèque, une analyse des pratiques discursives « à la Foucault ». Il nous paraît cependant possible de maintenir comme horizon théorique le système-archive établi par Foucault.

Le modèle proposé par Foucault, ici réduit à trois de ses dimensions, permet de dégager des éléments de nature à circonscrire les aspects du discours documentaire.

• *En tant qu'« espace limité de communication », polyphonique et historiquement réglé*, une formation discursive apparaît sous l'angle d'un « cadre interprétatif ».

Une formation discursive constitue une situation d'interprétation, un contexte pour un énoncé. Cette notion montre que l'on peut se passer des notions de thèmes, de genres textuels ou encore d'autorité d'auteurs pour relier entre eux des auteurs « qui ne se connaissent pas dans une trame dont ils ne sont pas maîtres ». Seul le maintien de la polyphonie, de l'hétérogénéité des textes, peut permettre l'émergence de domaines de savoir. Dans ce cadre, le texte isolé apparaît comme une unité « faible »⁹⁴ : il doit être pris dans un ensemble.

La notion de règles discursives qui délimite un espace de discours peut se poser comme alternative théorique aux pratiques actuelles de la sélection documentaire prises dans une impasse, comme nous le détaillerons ci-après : la sélection des sources recourt habituellement aux langages documentaires. Or cette intervention des langages documentaires à toutes les étapes du traitement des documents

⁹⁴ *Ibid.*, p. 34 : « L'unité matérielle du volume n'est-elle pas une unité faible, accessoire, au regard de l'unité discursive à laquelle il donne support ? ».

(sélection, indexation proprement dite ou encore exposition des documents) empêche de considérer les sources sous leur angle discursif et rend biaisé, peu clair, le choix effectué parmi les sources : sont sélectionnées uniquement les sources dont peut rendre compte le langage documentaire.

• *En tant qu'« archive »*, la formation discursive se donne sous la forme d'un système établissant, pour un énoncé donné, une relation entre son contexte de production et ses potentialités d'utilisation (cf. la relation énoncé-événement et énoncé-chose), cette relation établissant un niveau intermédiaire entre langue et corpus.

La notion de système-archive permet de donner consistance à notre hypothèse d'un discours documentaire : le discours documentaire se définit alors comme un niveau discursif spécifique dont la particularité est de maintenir les textes sous leur double forme d'événement et de chose. À ce titre, le discours documentaire où pourraient se déployer les systèmes de formations propres aux pratiques discursives peut se concevoir comme une transformation discursive réglée.

• *En tant que domaine de savoir*, la formation discursive repose sur les quatre dimensions suivantes : sémantique, énonciative, intertextuelle et fonctionnelle.

Cette approche des formations discursives met l'accent sur des aspects du discours documentaire qui nous étaient précédemment apparus en ordre dispersé. Les objets de discours, que peuvent construire les utilisateurs sur la base de descripteurs et qui peuvent être alors des « informations » pour eux, nécessitent l'inscription dans un domaine de savoir : l'indexation doit pouvoir spécifier les auteurs des sources et des documents en termes de « sujet » du discours, mais aussi établir des relations d'intertextualité et permettre un détournement des sources. L'émergence du savoir se réalise en effet à la marge des domaines constitués, par empiètement sur des textes dont l'appréhension première ne laisse pas forcément penser des formes de rapprochement, qu'en revanche une pratique discursive peut être amenée à établir.

Appréhendée sous trois facettes, la notion de formation discursive proposée par Foucault donne corps à la notion de discours documentaire comme espace d'organisation spécifique des documents, en montrant que l'on ne peut impunément s'en remettre aux textes tels qu'ils se donnent. Le modèle de Foucault oblige à penser, à expliciter les bases sur lesquelles s'effectuent l'exploration et le regroupement des sources en indexation : c'est d'abord cette contrainte de l'explicitation des principes de sélection qu'il nous semble important de retenir. La pratique d'indexation peut difficilement, nous semble-t-il, faire l'impasse sur cette question. En ce sens, il apparaît que, parmi les fondements théoriques de

l'indexation, doivent figurer des éléments touchant aux principes d'exploration des sources ; à ce titre, le modèle de Foucault fournit un cadre possible.

Par ailleurs, le modèle de Foucault insiste sur la nécessité de concevoir un énoncé sous l'angle de ses utilisations nouvelles. Il y a sur ce point aussi, dans le modèle de Foucault, des éléments pour un fondement théorique de l'indexation vue comme une opération de détournement d'usage des sources. Enfin, le modèle de Foucault nous permet de dégager une autre dimension susceptible, elle aussi, de participer aux fondements théoriques de l'indexation : c'est la nécessité de considérer les objets de l'indexation comme des objets discursifs, sous l'angle de leur épaisseur discursive.

Autrement dit, en posant dans l'horizon théorique de l'indexation la notion de système-archive ou de formation discursive au sens large, on est conduit à considérer l'indexation dans ses choix initiaux, qui sont d'ordre discursif, et non plus dans ses choix finaux, qui sont de nature lexicale : des documents ne peuvent être jugés semblables parce qu'affectés du même descripteur.

II.2- Le « système-archive » comme horizon théorique

En posant la notion foucauldienne d'archive à l'horizon de notre réflexion sur les fondements théoriques de l'indexation apparaissent des caractéristiques de la pratique documentaire rarement explicitées dans les discours de la pratique sur elle-même. En effet, mené à la lumière du modèle de Foucault, l'examen des pratiques d'indexation révèle, sous une forme parfois très ténue, les marques d'une « transformation discursive ». Pour faire apparaître cet aspect de l'indexation, nous poserons sous forme de conjecture que le processus de l'indexation relève d'une « fonction énonciative » (II.2.1). Nous tâcherons ensuite de mettre au jour les types de règles à l'œuvre dans le travail d'indexation compris dans sa phase de sélection des sources (II.2.2).

II.2.1- Conjecture : l'indexation comme « fonction énonciative »

Le processus de l'indexation partage avec le « système-archive » de Foucault le même objectif : faire voir un texte sous sa double face d'événement singulier (source) et de chose manipulable (document).

Pour atteindre cet objectif, l'indexation réalise une transformation qui porte, non sur le texte lui-même, mais sur son « entourage », son contexte ; en ce sens, on peut la rapprocher de la notion de « fonction énonciative » proposée par Foucault, fonction qui permet à un texte d'exister comme un « énoncé ». Le document pourra être vu, en indexation, comme un énoncé dans ce cadre-là et l'indexation comme la fonction d'existence d'une source comme document.

La fonction énonciative est approchée par Foucault dans les termes suivants :

La fonction énonciative - montrant bien par là qu'elle n'est pas pure et simple construction d'éléments préalables - ne peut s'exercer sur une phrase ou une proposition à l'état libre. Il ne suffit pas de dire une phrase, il ne suffit même pas de la dire dans un rapport déterminé à un champ d'objets ou dans un rapport déterminé à un sujet, pour qu'il y ait énoncé - pour qu'il s'agisse d'un énoncé : il faut la mettre en rapport avec tout un champ adjacent. [...] Un énoncé a toujours des marges peuplées d'autres énoncés. [...] Il n'y a pas d'énoncé en général, d'énoncé libre, neutre et indépendant mais toujours un énoncé faisant partie d'une série ou d'un ensemble, jouant un rôle au milieu des autres, s'appuyant sur eux et se distinguant d'eux, il s'intègre à un jeu énonciatif, où il a sa part aussi légère, aussi infime, qu'elle soit. Foucault, 1969, respectivement p. 128 et p. 130.

On peut observer, dans les pratiques d'indexation, les traces de l'exercice de l'indexation comme fonction énonciative :

- (i) chaque pratique d'indexation établit une organisation spécifique des documents. Indexé par un même mot, un document n'est pas pour autant introduit dans la même « série » pour reprendre les termes de Foucault : le « champ adjacent » retenu pour un énoncé-document n'est pas toujours le même. Il y a là, nous semble-t-il, la marque d'une fonction énonciative : l'indexation se signale par la série des documents qu'elle constitue (II.2.1.1) ;
- (ii) les pratiques d'indexation utilisent divers procédés pour maintenir l'aspect dual du texte (source et document) : on observe des moyens pour maintenir la trace des conditions de production d'une source ; on relève aussi les marques d'un « jeu énonciatif » qui permet à un document-énoncé de trouver sa place dans une série. C'est en ce sens que le document pourra apparaître comme un « énoncé » au sens foucauldien du terme (II.2.1.2).

En affaiblissant les concepts établis par Foucault, nous proposons ci-après une analyse des pratiques d'indexation qui permet, sinon de définir à proprement parler l'indexation comme une fonction énonciative, du moins d'exhiber des mécanismes discursifs, tenus le plus souvent hors du cadre d'appréhension de l'indexation, comme le fait remarquer Suzanne Bertrand-Gastaldi :

En somme, les « traductions successives » des textes scientifiques par l'homme et la machine font progressivement disparaître le discours au profit du vocabulaire, la syntaxe et les indices au profit des symboles. Alors que le discours est une actualisation de la langue, il est presque retransformé, dans les bases de données, en langue artificielle, reconstituable à partir des formations discursives. Il reste cependant des traces irréductibles des conditions d'énonciation.
Bertrand-Gastaldi, 1989, p. 22.

Si nous rejoignons Suzanne Bertrand-Gastaldi sur la nécessité d'une reformulation de l'indexation dans des termes qui fassent voir sa dimension discursive, nous ne pensons pas que, en l'état actuel, les bases de données et les types d'accès qui s'y trouvent puissent permettre une reconstitution aisée des « formations discursives » : nous y reviendrons.

II.2.1.1- L'organisation des discours en indexation

Si, comme nous l'avons précédemment mentionné, la littérature courante sur l'indexation relève régulièrement la variabilité des termes utilisés pour indexer les mêmes textes, elle relève moins la variabilité des textes qui se trouve sous l'indexation par un même terme.

Pour illustrer cet aspect, nous prendrons d'abord un exemple extrait de l'expérimentation que nous avons menée⁹⁵.

Le même descripteur, « Arte », est utilisé par six organismes documentaires pour indexer un même document issu du journal *Le Monde*⁹⁶. Or, ce même descripteur ne renvoie pas toujours aux mêmes textes. On retrouve ici la variabilité de mise en documents que nous avons précédemment relevée, mais le point consiste ici à remarquer que cette variabilité n'est pas exprimée par le biais du descripteur : la variabilité des mises en discours reste invisible aux yeux d'un utilisateur.

Ainsi, si l'on imagine la requête suivante adressée aux six organismes documentaires analysés - « Dans *Le Monde* du 1er décembre 1994, qu'y a-t-il eu sur Arte ? » -, on obtiendra trois types de réponse différents :

- (i) un article : pour quatre des six organismes documentaires, seul l'un des deux articles du *Monde* est affecté du terme « Arte » ;
- (ii) deux articles : pour un centre de documentation, les deux articles sont chacun indexés par le terme « Arte » ;
- (iii) quatre articles : l'un des six organismes documentaires a intégré les deux articles du *Monde* dans une revue de presse comportant des articles issus d'autres sources.

⁹⁵. L'expérimentation est présentée en annexe 2.

⁹⁶. *Le Monde* du 1er décembre 1994 propose en page 10 deux articles : le premier s'intitule « ARTE veut élargir son public et casser son image de chaîne élitiste », le second, un encadré, porte le titre « Jean-Marie Cavada, premier président du GIE ».

Encore ne mesurons-nous ici la variabilité de mise en discours uniquement en considérant une seule source d'information.

En interrogeant les catalogues de bibliothèque qui utilisent le même langage documentaire, le langage Rameau, on retrouve la même variabilité de mise en discours sous l'emploi d'un même terme. Ainsi, à l'interrogation par le terme « analyse documentaire », on trouve le même ouvrage (Chaumier 1988) mais alors qu'il se situe au sein d'une série de 22 éléments dans la bibliothèque de l'ENSSIB, il est inscrit dans une autre série de 6 éléments dans le catalogue du réseau BRISE à Saint-Étienne. Certes, les deux bibliothèques ont procédé à la même sélection de source qu'elles présentent par le même accès, mais l'inscription de l'ouvrage dans un ensemble est radicalement différente. Cette différence traduit une variation dans les principes de sélection des ouvrages retenus, variation qui n'est pas rendue visible aux utilisateurs alors qu'elle constitue, nous semble-t-il, la marque même de l'indexation, voire tout son intérêt. Une indexation doit permettre de fournir un regard problématisé sur la production éditoriale ; or si problématisation il y a, sous la forme de règles de sélection, elle est maintenue inaccessible.

Ces deux exemples montrent que l'indexation réalise une organisation des discours spécifique, qui présente la singularité de ne laisser aucune « trace » visible ; ce n'est pas pour autant qu'elle ne suit aucune règle. En examinant les éléments qui constituent le contexte d'un document, on peut relever des « traces » qui indiquent que la « fonction énonciative » réalisée en indexation, si elle n'est pas entièrement réglée, est du moins partiellement régie par les sources elles-mêmes.

II.2.1.2- Traces de décontextualisation et de recontextualisation dans les objets documentaires

À la fin du chapitre III, nous avons proposé de considérer l'indexation comme une opération de contextualisation réalisée en deux temps : à une décontextualisation des sources succède une recontextualisation au terme de laquelle une source devient un document. Cette première approche de l'indexation trouve un fondement théorique dans la notion de système-archive proposée par Foucault, qui établit précisément un niveau où les deux faces d'un texte - événement et chose - peuvent être captées : c'est, dans le modèle de Foucault, le niveau où se forment les objets discursifs propres à une pratique discursive.

S'il est clair que les pratiques d'indexation sont encore loin de constituer, dans leur

espace de discours, ce niveau des formations discursives, il semble qu'elles mettent en œuvre des procédés pour maintenir un texte sous son double aspect de source et de document.

A- Marques de la source dans le document : traces de la décontextualisation

Dans la chaîne de traitement documentaire telle qu'elle est traditionnellement présentée dans les manuels ou les traités, l'opération d'indexation est distinguée de celle du catalogage (ou description bibliographique) qui consiste à prélever, de l'entourage d'une source, des données que l'on nomme factuelles et qui comprennent des mentions relatives aux titres, auteurs, éditions, dates, pagination, langue, etc.

Dans le cadre d'une approche de l'indexation en termes de transformation discursive, cette formalisation de données externes au « contenu » du texte participe à la création du document. Elle correspond, nous semble-t-il, à une tentative pour maintenir, dans le cadre du document, quelques-unes des particularités du contexte de production d'une source. Cette présence des éléments de la source au sein du document n'est pas sans incidence sur l'interprétation des termes d'indexation, comme le remarque Suzanne Bertrand-Gastaldi :

Deux indexations en tout point semblables, de par leur rattachement à deux documents différents publiés dans des revues différentes, à des dates différentes, constitueront bien deux énoncés différents qu'il faudra bien interpréter différemment avec l'aide du co-texte et de l'intertexte qui, lui, contient donc aussi un certain nombre d'éléments indiciels.

Bertrand-Gastaldi, 1989, p. 12.

En ce sens, il nous semble que les pratiques documentaires parviennent, dans une certaine mesure, par le biais de la description bibliographique, à maintenir l'hétérogénéité des sources.

Sur ce point, il nous semble que le document doit être en indexation défini comme un objet complexe, comprenant, outre le texte de la source et les descripteurs, les données factuelles issues de la description bibliographique.

Appréhendé sous cet angle, un document peut être compris comme un énoncé, du moins dans le sens foucauldien du terme. En effet, selon Foucault, il n'est d'énoncé que situé, et situé doublement : par rapport à son contexte de production d'une part⁹⁷ et par rapport à son contexte d'utilisation d'autre part⁹⁸.

⁹⁷. Foucault 1969, p. 13 : « Les coordonnées et le statut matériel de l'énoncé font partie de ses caractéristiques intrinsèques. [...] Il faut qu'un énoncé ait une substance, un support, un lieu et une date. Et quand ces conditions requises se modifient, il change lui-même d'identité ».

⁹⁸. *Ibid.*, p. 130 : « Il n'y a pas d'énoncé en général, d'énoncé libre, neutre et indépendant mais toujours un énoncé faisant partie d'une série ou d'un ensemble, jouant un rôle au milieu des autres,

En ce sens, le document, si on l'entend comme un texte pourvu de ses « coordonnées », peut fonctionner comme un énoncé puisqu'il permet d'établir ainsi un lien avec la source d'où il provient. En effet, comme le précise encore Foucault, « une série de signes deviendra un énoncé à condition qu'elle ait à “ autre chose ” un rapport spécifique qui la concerne elle-même, et non point la cause, non point ses éléments »⁹⁹. Tel que nous l'avons envisagé jusqu'à présent, le document nous paraît être un énoncé dans la mesure où il entretient une relation avec « autre chose », ce que nous avons nommé la « source », cette relation ne relevant ni de la causalité ni du définitionnel, mais de l'interprétation : nous avons en effet posé le document comme étant l'interprétant de la source¹⁰⁰.

Ouvrons ici une parenthèse qui soulignera à nouveau le rôle opacifiant du langage documentaire sur la perception des sources en indexation.

Le document peut permettre l'interprétation d'une source grâce aux autres documents auquel il y est lié *via* les descripteurs. Sur ce point, les descripteurs sont considérés, dans l'approche de l'indexation que nous proposons, comme des marques de l'insertion d'un document dans une série plus que comme des marques de son contenu propre. L'approche classique des langages documentaires, et des thésaurus en particulier, tend à obscurcir ce phénomène. En effet, selon les définitions normatives, la particularité des thésaurus tient aux relations qui y sont établies entre les termes¹⁰¹ (relations de généralité, de spécificité, d'équivalence ou d'association) : c'est la stabilité de ces relations qui est posée comme garante d'une « bonne » indexation. Or, comme le souligne Suzanne Bertrand-Gastaldi, les relations qu'entretiennent les descripteurs au sein d'une base de données ne sont plus les mêmes que celles proposées dans le thésaurus¹⁰². Rien de plus normal puisque le corpus qui a permis d'établir le thésaurus n'est pas le même que celui qui est soumis à l'indexation.

B- Marques du document : traces de la recontextualisation

La recontextualisation s'exprime prioritairement par le biais des descripteurs qui signalent l'inscription d'une source dans un nouvel ensemble où vont pouvoir se

s'appuyant sur eux et se distinguant d'eux ».

⁹⁹. *Ibid.*, p. 117.

¹⁰⁰. Voir précédemment chapitre III § III.1.

¹⁰¹. Cf. norme Z 47-100 (1981), p. 7 : « Une des fonctions primordiales d'un thésaurus est de représenter les relations entre concepts par l'indication des rapports entre les termes utilisés pour les décrire. Le réseau des relations d'un descripteur avec les autres termes (descripteurs ou non-descripteurs) fournit ainsi une sorte de définition et concourt à réduire les risques d'ambiguïté en situant le descripteur dans un contexte qui en précise le sens ».

¹⁰². Bertrand-Gastaldi 1989, p. 12 : « Chaque mot-clé attribué à un texte entretient avec les autres des relations syntagmatiques et l'ensemble des mots-clés attribués dans une banque de données n'a rien à voir avec l'ensemble des descripteurs d'un thésaurus et de leurs relations paradigmatiques ».

déployer de nouvelles utilisations, de nouvelles possibilités d'interprétation :

L'insertion d'un texte sous une rubrique de classification, d'un descripteur ou d'un ensemble de descripteurs par un indexeur donne un nouvel éclairage à ce texte, crée un nouveau réseau de relations en le rapprochant à la fois du système de classification ou du thésaurus et de l'ensemble des autres textes qui ont déjà reçu les mêmes « étiquettes ».

Bertrand-Gastaldi 1993, p. 145.

Dans ce cadre, le nouveau contexte d'un document est constitué par l'ensemble des autres documents regroupés sous le même descripteur. C'est à ce stade que la source acquiert une existence spécifique, une « existence documentaire » (une existence en tant que document), qui nous paraît être le résultat spécifique de la fonction « indexation », sa spécificité tenant principalement dans le « détournement d'usage » qu'elle met en place. Il importe à nouveau de rappeler cette évidence qu'il n'existe aucune source explicitement dédiée à l'indexation et que seule une fonction de « transformation » peut permettre de créer un document.

Si la marque la plus visible de la « fonction énonciative » par laquelle nous nous proposons de décrire l'indexation reste les descripteurs, d'autres aspects se révèlent au niveau spécifique de l'espace documentaire. En effet, si la source est inscrite dans le document sous la forme de données factuelles, le document est aussi inscrit dans le choix des sources : soit sous la forme directe d'une relation d'intertextualité entre sources à sélectionner et documents déjà constitués, soit sous la forme indirecte du thésaurus, où le poids du corpus présidant à sa création influence le choix des sources retenues. Nous abordons ce dernier aspect ci-après dans le cadre de la problématique plus générale des principes d'exploration des sources en indexation.

Comme dans les paragraphes précédents, où l'on a mené une analyse des pratiques d'indexation sous l'angle de la notion de « fonction énonciative », c'est l'éclairage fourni par la notion de « système-archive » établie par Foucault qui nous permettra de problématiser les règles de sélection en indexation.

II.2.2- Problématique des règles d'exploration des sources en indexation

La question des règles orientant le choix des sources et légitimant leur existence en tant que documents, que ce soit en bibliothèque ou dans les centres de documentation, constitue une préoccupation récente dans le milieu professionnel,

comme le note Bertrand Calenge¹⁰³, auteur d'un des rares ouvrages français sur les « politiques d'acquisition ». À l'heure actuelle, les structures documentaires pourvues de « chartes » explicitant les critères de choix restent beaucoup moins nombreuses que celles qui en sont dépourvues. Un véritable « tabou » pèse sur le processus d'acquisition, remarque Calenge¹⁰⁴.

L'enjeu est pourtant de taille. Les bibliothèques et les centres de documentation abondent d'anecdotes relatant le poids des affinités électives d'un individu pour un domaine ou un auteur, créant à terme des trous béants dans les collections documentaires comme des sur-représentations inexplicables de certains sujets ; de quoi rendre incohérent le résultat de requêtes documentaires, la meilleure indexation possible ne pouvant guère pallier le déséquilibre des collections elles-mêmes. De façon moins anecdotique, une enquête menée auprès des publics de la Bibliothèque nationale¹⁰⁵ a montré que la faible fréquentation des chercheurs dans le domaine dit des sciences « dures » est directement liée à une absence de collections les concernant¹⁰⁶.

Ces dernières années, plusieurs facteurs ont contribué à la création de chartes destinées à expliciter les règles d'exploration des sources, notamment :

- les impératifs budgétaires qui frappent tous les établissements culturels ont définitivement mis fin à toute velléité d'exhaustivité ;
- la notion de réseau documentaire a permis aux structures documentaires de penser les acquisitions en termes de complémentarité, tandis qu'elle exigeait une définition claire des champs de connaissances investis et des critères de sélection ;
- la multiplicité des types de structures documentaires, et notamment l'émergence des bibliothèques de lecture publique, a modifié la perception des bibliothèques et de leurs fonds. Il ne s'agit plus d'accumuler des documents, dont la pertinence sera bien découverte, un jour, par un chercheur, mais de présenter une collection cohérente susceptible de fournir des orientations claires à ceux que la sociologie des publics en bibliothèque nomme les « faibles lecteurs » ;
- plus globalement, il est apparu aussi et essentiellement que la notion de collection documentaire et celle de politique d'acquisition qui la

¹⁰³. Calenge 1994.

¹⁰⁴. *Ibid.*, p. 24.

¹⁰⁵. Étude non publiée menée sous la direction de Christian Baudelot. Une synthèse se trouve dans Baudelot et Véry 1994.

¹⁰⁶. En effet, pendant près de trente ans, la Bibliothèque nationale a délaissé l'acquisition de documents en langue étrangère dans le domaine des sciences « dures ».

sous-tend constituaient en propre le métier des professionnels de l'information et de la documentation¹⁰⁷.

On dispose désormais, pour une part, de règles explicites d'exploration des sources : on synthétisera celles que propose Calenge. On présentera aussi celles qui nous ont été fournies au cours de notre enquête, qui concernent plus spécifiquement les critères qui, pour une source donnée, guident le choix des segments textuels à indexer.

Reste une part importante de règles implicites, que des chercheurs en sciences de l'information ont fait émerger : nous joindrons à leurs remarques celles que nous avons pu faire au cours de notre expérience.

II.2.2.1- Les règles explicites d'exploration des sources

Qu'il s'agisse de sélection de sources (A) ou de sélection d'objets textuels à indexer (B), c'est le recours à un référentiel existant qui est le plus souvent utilisé par les professionnels. Ces référentiels sont constitués des grands domaines de connaissance ou d'une liste de thèmes jugés spécifiques au public d'un établissement. C'est sur ce point que l'on retrouve les problèmes pointés par Foucault, les *a priori* formels prenant le pas sur les *a priori* historiques.

A- Outils utilisés pour la sélection des sources

Calenge [1994] propose¹⁰⁸ deux types d'outil pour guider les acquéreurs dans leur tâche de sélection des sources : d'une part, la classification décimale Dewey, d'autre part, une échelle (à établir par chaque établissement) codant les usages possibles d'un document (en fonction de leur « complexité »)¹⁰⁹.

Sans revenir ici sur les critiques dont a pu faire l'objet la classification Dewey en matière de représentation des domaines de connaissance, on peut s'étonner de voir intervenir, dès le choix des sources, un langage documentaire.

Trois types d'arguments sont avancés par Calenge. Le premier tend à justifier la

¹⁰⁷. Calenge 1994, p. 401 : « Gérer, développer et renouveler une collection est l'objet même du métier de bibliothécaire : toutes les sciences appelées à la rescousse (sociologie, linguistique, psychologie...), toutes les techniques extérieures (architecture, informatique...) ou développées de façon interne (catalogage, indexation...), tournent autour de cet objet unique qui légitime le bibliothécaire et constitue le cœur de son activité : la collection vivante et en action ».

¹⁰⁸. Calenge s'appuie pour ce faire sur l'analyse de chartes documentaires existantes.

¹⁰⁹. Ce second outil (évaluant les « niveaux de lecture » d'un document : adulte, enfant, etc.) n'est pas utilisé pour sélectionner à proprement parler les sources documentaires ; il opère plutôt un filtrage sur un ensemble de sources sélectionnées à partir de la classification Dewey. C'est pourquoi nous n'aborderons pas ici ce deuxième type d'outil.

classification Dewey au regard de l'utilisation d'un autre langage documentaire¹¹⁰. Les deux autres mettent en avant des aspects pragmatiques : la classification Dewey constitue le « langage commun » des professionnels¹¹¹ ; elle est par ailleurs utilisée dans les catalogues des éditeurs comme par les intermédiaires qui diffusent les nouveautés éditoriales¹¹².

Le choix d'un langage documentaire pour trier parmi les sources existantes n'est donc pas à proprement parler problématisé :

- on ne tient pas compte du fait que la classification Dewey, si elle « synthétise le contenu » d'un document, n'est pas un outil neutre. Autrement dit, les stratégies de choix propres aux professionnels qui semblaient être mises en avant par la notion de « politique d'acquisition », sont ici réduites au silence : on s'en remet à la notoriété d'un instrument comme si les bibliothécaires n'en avaient pas été les agents actifs ;
- on ne distingue pas les deux stratégies d'exploration des sources et d'exposition des documents : en recourant au même outil pour trier les sources et pour classer les documents, ce qui devait constituer en propre le métier du bibliothécaire, la mise en collection des documents, ne trouve plus ici de moyens spécifiques.

On le voit : les règles d'exploration des sources, telles qu'elles sont habituellement établies en bibliothèques, ne sortent pas d'une approche lexicale du contenu. La même perception des mots et des choses qui caractérise, dans la démarche classique, l'indexation d'un document¹¹³, se retrouve, sous une forme identique, lorsqu'il s'agit de formaliser les critères de choix d'une source. La distinction source/document n'est pas établie ni les stratégies dévoilées.

Ne sont pas non plus prises en compte les particularités des langages documentaires. Comme l'ont établi Bourion et Malrieu [1994], sur la base d'une étude d'un plan de classification en psychologie sociale, on ne saurait ignorer les « instructions de lecture » implicitement à l'œuvre dans un langage documentaire. Nous renvoyons, pour le détail, à l'analyse¹¹⁴ qu'elles mènent pour exhiber ce qu'elles nomment le « discours classificatoire »¹¹⁵ des indexeurs, en adoptant une

¹¹⁰. Voir Calenge 1994, p. 129 : « C'est le principe décimal des classifications qui reste un atout dans la définition et l'analyse des objectifs documentaires car il autorise des regroupements exhaustifs interdits aux descripteurs alphabétiques (liste Rameau ou thésaurus spécialisés) ».

¹¹¹. *Ibid.*, p. 124 ou p. 134, entre autres.

¹¹². *Ibid.*, p. 396.

¹¹³. Voir notre première partie : les problèmes théoriques de l'indexation.

¹¹⁴. Bourion et Malrieu 1994, p. 83-131 : elles étudient le plan de classement en tant que « genre textuel » de façon à cerner les contraintes qu'il fait peser dans la construction de l'interprétation.

¹¹⁵. Nous y revenons ci-après, § III.2.1.

approche à la fois cognitive et linguistique. Les résultats de leur analyse rejoignent ceux de Sylvie Bruxelles, précédemment évoqués¹¹⁶, sur les relations d'intertextualité qui s'établissent entre plans de classement et textes à indexer. Ce que ces auteurs montrent des contraintes de lectures exercées par les langages documentaires sur le choix des mots nous semble être tout aussi actif dans le choix des sources.

En outre, il importe de rappeler ce qui est régulièrement observé par les professionnels eux-mêmes. Le rangement d'une source sous un indice de classification est toujours le résultat d'une distorsion, plus ou moins forte, et au final toujours le résultat d'un choix¹¹⁷ : les sources se classent difficilement elles-mêmes dans le cadre d'une représentation des connaissances établie sur des bases depuis longtemps perdues et oubliées¹¹⁸, modifiée de façon *ad hoc* au gré des difficultés rencontrées. De là, un certain nombre de « licences » que s'accordent nécessairement les documentalistes et les bibliothécaires, sans toujours justifier ou simplement expliciter les libertés prises par rapport au référentiel de base.

Un exemple parmi les nombreux cas que l'on trouve dans les pratiques documentaires est relevé par Bourion et Malrieu¹¹⁹ : le CNRS dispose de deux bases de données qui distinguent, d'une part, « les sciences humaines, sociales et économiques » (base Francis) et, d'autre part, « les sciences, les technologies et la médecine » (base Pascal). Il faut connaître les péripéties de l'histoire de l'institution pour savoir que le domaine de la psychologie ne fait pas partie, dans les bases de données du CNRS, du domaine des sciences humaines mais de celui des sciences de la vie (base Pascal).

Là encore, ce qui vaut pour les pratiques de classement vaut aussi pour les pratiques d'exploration des sources, puisque ce sont les mêmes outils qui sont employés. Dans le cadre d'une organisation du travail en général très parcellisée comme il est courant dans les bibliothèques d'importance, un acquéreur n'a de vue sur son domaine que ce que lui propose un indice Dewey : sa stratégie d'exploration se trouve de fait très limitée. C'est ainsi que si d'aventure au sein d'une base de données, deux documents relevant d'indices Dewey différents venaient à être rapprochés, ce serait là un pur hasard, un fait de rencontre qui ne devrait rien à une

116. Bruxelles 1991, voir chapitre II § III.1.

117. Bourion et Malrieu 1994, p. 86 : « Mais les " objets " à classer (les articles) renvoient à des concepts qui peuvent nécessiter la référence à différentes spécialités, se situer à la périphérie des courants dominants parce qu'ils s'autodéfinissent au travers de concepts et " territoires nouveaux ", en émergence ».

118. Ainsi de la classification décimale Dewey, établie en 1873 sur la base de la littérature alors disponible dans les bibliothèques américaines d'Amherst, des régions de New-York et New England. Un historique peut être trouvé dans Comaroni 1988 par exemple.

119. Bourion et Malrieu 1994, p. 86.

stratégie de regroupement. C'est en ce sens qu'il paraît très délicat de pouvoir *a posteriori*, comme le suggérait Suzanne Bertrand-Gastaldi, rétablir, par le biais de « formations discursives », des connexions entre documents. Selon nous, il importe de desserrer, dès le niveau de la sélection documentaire, l'étau des langages documentaires.

B- Outils utilisés pour la sélection des objets à indexer

Dans le cadre de notre expérimentation (où l'on cherchait à définir ce qui guidait la sélection des objets d'indexation au sein d'une source considérée en soi comme pertinente), si l'on retrouve le même recours à des référentiels établis sur des critères de contenu, on trouve aussi d'autres types de règles.

De façon générale, les organismes documentaires étudiés s'appuient, pour établir leur sélection, soit sur une liste de thèmes (par exemple : la communication, les arts, les médias, la vie culturelle, etc.), soit sur les rubriques du journal *Le Monde* : pour un organisme documentaire spécialisée dans l'économie par exemple, seules les pages ainsi intitulées seront explorées.

Ce recours à un mode de guidage émanant de la source elle-même peut être, dans certains cas, problématique. Sur ce point, un entretien mené sur la sélection documentaire exercée par un documentaliste utilisant les règles d'exploration par rubriques nous a montré le caractère réducteur de ce type d'exploration. Habitué à ne consulter que les pages « Communication » du journal, le documentaliste n'avait pas retenu un article qui, classé dans la rubrique « Économie », aborde le domaine des technologies nouvelles habituellement traité par le service de documentation : l'article en question¹²⁰ (« La Route intelligente ») rend compte du premier congrès sur la télématique des transports. Le documentaliste, interrogé sur la non-sélection de cet article, reconnaissait la pertinence de cet article pour les domaines qu'il couvre, tout en hésitant à le retenir à cause de son apparition dans la rubrique « Économie ».

Sur ce point, le système des rubriques dans le journal *Le Monde* est perçu de la même façon que le système des indices de la classification Dewey : les deux sont appréhendés comme des référentiels objectifs et nécessairement valides, sans que la particularité de leur construction ne soit prise en compte. Si nous ne remettons pas nécessairement en cause le recours à des référentiels établis, nous remettons en cause la perception dont ils font l'objet. Il semble en effet nécessaire de les considérer comme porteurs d'« instructions de lecture », pour reprendre les termes de Bourion et Malrieu, qui peuvent intéresser le lecteur à ce titre-là : c'est alors le

¹²⁰. *Le Monde*, 1/12/1994, p. 19.

point de vue du *Monde* sur les événements qui peut être retenu et indexé comme tel et non les événements en eux-mêmes.

Sur ce point se distinguent, parmi les organismes documentaires consultés, deux types de perception différente qui déterminent deux types de stratégie :

- (i) pour ceux qui considèrent *Le Monde* comme une source d'information comme une autre, le quotidien sera lu au travers d'un référentiel basé sur le contenu (liste thématique externe ou système de rubriques interne au journal) ;
- (ii) pour ceux qui considèrent *Le Monde* comme source spécifique dans la presse française, les stratégies d'exploration seront moins déterminées par un contenu supposé : les signatures des articles pourront, par exemple, prendre plus d'importance que le contenu traité. Perçu sous l'angle du « journal de référence » (l'expression revient souvent dans les entretiens), le journal, dont il semble important d'avoir le point de vue, sera alors exploré de façon large et selon des règles qui restent alors très implicites.

Il nous semble que l'on peut voir, dans ce dernier type d'exploration qui cherche à prendre en compte la spécificité de la source, les prémices d'une démarche propre à intégrer la notion de formation discursive établie par Foucault.

De la même façon que, précédemment dans le point A, apparaissaient, au-delà de l'adoption de règles explicites, certaines zones de liberté prises par les acquéreurs (par rapport aux plans de classement), se dessinent également, dans le cadre de notre expérimentation, quelques décalages entre principes de sélection avoués et principes de sélection appliqués.

Ainsi, le secteur « presse » de *La Documentation française* met-il en avant, dans ses déclarations de principe, « la pluralité des sources [qui] apporte à l'utilisateur une information équilibrée, complète et impartiale »¹²¹. En effet, le secteur « presse » de *La Documentation française* est l'un des rares services d'information publics qui traitent la presse de tous les partis politiques (De *Présent à Révolution* en passant par *Rouge et Vert*) comme de tous les syndicats (*FO hebdo*, *CFDT magazine*, etc.). À partir de ces sources, les critères de sélection explicites sont les suivants¹²² :

- « exhaustivité sur l'organisation et l'activité des pouvoirs publics ;
- large sélection sur la vie des partis, des syndicats, du patronat ;
- sélection rigoureuse quant au contenu informatif sur les questions

¹²¹. *Infos-Bipa*, mise à jour 1995, [p. 1].

¹²². *Ibid.*, [p. 2].

d'actualité liées au débat politique (problèmes de société, grandes entreprises, vie des médias, nouvelles technologies, etc.) ».

On devrait logiquement s'attendre à ce que la sélection exercée sur *Le Monde* soit finalement assez « rigoureuse » puisque sont abordées essentiellement dans ce journal des « questions d'actualité liées au débat politique ». Or, pour l'année 1994, les références proposées aux utilisateurs sont extraites pour un tiers (26%) du *Monde*, qui arrive en tête du palmarès des sources présentées dans la base de données de *La Documentation française*. À eux trois, les quotidiens *Le Monde*, *Le Figaro* et *Libération* représentent, en 1994, 60% des articles proposés pour rendre compte de la vie politique française. Il n'y a là, sans doute, rien de très étonnant quand on connaît le poids de chacun de ces quotidiens dans le paysage de la presse française ; reste que les règles explicites de sélection ne sont pas celles qui sont appliquées.

Une analyse de type sociologique pourrait sans doute contribuer à faire émerger les règles implicites effectivement appliquées.

Sous l'angle de vue que nous avons retenu, on essaiera de dégager les principes de sélection qui peuvent être mis en œuvre, en reprenant les études que d'autres chercheurs ont pu mener sur un plan plus linguistique.

II.2.2.2- Les principes implicites d'exploration des sources

À la suite de Suzanne Bertrand-Gastaldi [1989], qui reprend elle-même les conclusions de chercheurs anglophones comme Beghtol [1986], on peut dégager un principe général d'exploration des sources en indexation : le principe de l'intertextualité, actif à plusieurs niveaux que nous indiquerons succinctement.

Les théories de l'intertextualité sont nombreuses, diverses, souvent prises en défaut pour celles qui insistent sur le rapprochement entre textes effectué sur la base du seul contenu¹²³. Sans pouvoir entrer dans les problématiques spécifiques à ces théories, nous proposons de considérer l'intertextualité sous un angle qui insiste sur la transformation des usages d'un texte, lorsqu'il passe d'un ensemble de textes à un autre. C'est pourquoi nous retiendrons l'approche suivante de l'intertextualité :

[L'intertextualité] absorbe l'énoncé qu'elle emprunte à un modèle antérieur pour l'inscrire dans un autre ensemble textuel : elle ne se contente pas toutefois de l'incorporer, elle le soumet à une activité transformatrice, elle enchâsse le texte primitif dans un contexte nouveau dans le dessein d'en modifier le sens. L'intertextualité ne recouvre ainsi pas uniquement une opération mémoriale et

¹²³. Marandin [1979] et [1993] ainsi que Pêcheux [1975] ont pu mener la critique sous cet angle.

assimilatrice, elle n'est pas uniquement une transplantation d'un texte dans un autre, mais elle se définit par un travail d'appropriation et de réécriture qui s'applique à recréer le sens, en invitant à une lecture nouvelle.

Eigeldinger, 1987, p. 11, cité in Bertrand-Gastaldi, 1993, p. 145.

Le principe de l'intertextualité ainsi compris met en valeur ce que nous avons appelé l'espace discursif propre à l'indexation, l'espace d'interprétation des sources, réalisé bien avant l'assignation ou l'extraction effectives de descripteurs.

L'intertextualité, la mise en rapport de textes dans le but de « construire un univers relationnel »¹²⁴, se réalise sur plusieurs plans dans la phase de sélection des sources :

A- entre les sources elles-mêmes ;

B- entre les documents déjà sélectionnés et les sources explorées ;

C- entre les usages antérieurs des sources et les sources explorées.

A- Type 1 d'intertextualité : au niveau des sources

Cet aspect de l'intertextualité est surtout visible dans les cas de sélection d'objets d'indexation parmi un ensemble de sources finies (c'est le cas de notre expérimentation qui porte sur le dépouillement d'une source retenue *a priori*). La décision qui préside au choix de tel ou tel article du *Monde* comme au choix de son insertion dans une série (revue de presse, par exemple) se prend au regard des autres sources (d'autres quotidiens par exemple), c'est-à-dire sur la base d'un jugement qui repose sur le « déjà-dit », sur la base d'un contenu jugé semblable. De là deux types de stratégie : soit le « déjà-dit » apparaît comme une redondance et la source consultée n'est pas retenue ; soit il apparaît au contraire comme présentant un point de vue complémentaire et la source est alors sélectionnée à ce titre. Les deux cas de figure se trouvent dans notre enquête. À partir d'un même jugement de ressemblance (notion de « déjà-dit »), deux types de stratégies différentes sont adoptées sur la base de règles qui restent implicites.

B- Type 2 d'intertextualité : au niveau des documents

Comme nous l'avons précédemment évoqué, les documents déjà constitués pèsent sur le choix des sources, soit de façon directe soit de façon indirecte quand la sélection des sources se mène par le biais d'un langage documentaire : comme le note Bertrand-Gastaldi, « l'interprétation des nouveaux textes subit donc en partie l'influence des textes antérieurs »¹²⁵. Là encore, si c'est un même type d'intertextualité qui est mis en œuvre, les décisions prises peuvent être opposées.

¹²⁴. Ricardou cité in Bertrand-Gastaldi 1989, p. 142.

¹²⁵. Bertrand-Gastaldi 1989, p. 147.

Dans certains cas, seules les sources présentant une « nouveauté » par rapport à l'existant seront constituées comme documents ; dans d'autres cas, seules les sources faisant « suite » aux documents en place seront retenues. Là encore les règles qui président à ces choix ne sont pas explicitées. À l'analyse, apparaissent des stratégies qui, pour être établies sur le même constat, aboutissent pourtant à des décisions opposées.

C- Type 3 d'intertextualité : au niveau des usages antérieurs

L'utilisation faite des sources dans des contextes qui peuvent être extra-documentaires n'est pas non plus sans exercer une influence sur la sélection ou le rejet d'une source : là encore, les décisions prises à partir de constats semblables peuvent être opposées et les stratégies restent non dévoilées.

Odile Le Guern, travaillant sur le traitement documentaire de l'image, relève que le droit d'accès d'une image dans une iconothèque peut être contraint par des usages antérieurs :

Il est parfois nécessaire de tenir compte du contexte que constitue le réseau des utilisations précédentes, des précédents discours qui ont intégré les documents. [...] Certaines images sont devenues des images symboles du premier discours qui les a intégrées et ne pourront que très difficilement faire l'objet d'autres lectures pour d'autres utilisateurs, originales par rapport à ce premier discours.
Odile Le Guern, 1989, p. 428.

Si certains usages bloquent la transformation d'une source en document, inversement, les usages extra-documentaires d'une source peuvent décider de son intégration *a posteriori* dans l'espace documentaire. C'est le cas en particulier pour les indexeurs travaillant sur la presse d'actualité, l'entrefilet d'un jour pouvant, quelques semaines plus tard, se lire comme le début de toute une chaîne d'événements, méritant alors d'être intégré dans la collection documentaire.

L'existence documentaire d'une source n'est donc pas intrinsèquement liée à un type de texte. C'est ainsi que tous les organismes documentaires interrogés dans notre enquête, sans exception aucune, gardent une collection complète du journal *Le Monde*, au moins sur une année, malgré le travail de dépouillement, parfois très long (pour certains, près de deux heures consacrées uniquement à la sélection) dont il a fait l'objet. Certains d'entre eux¹²⁶ souscrivent parallèlement un abonnement à la base de données du *Monde*, pour éventuellement récupérer des articles anciens qui n'auraient été pas retenus. En somme, on retrouve, malgré tous les discours sur la valeur d'échange des réseaux, sur l'impossibilité de l'exhaustivité ou, plus positivement, sur la nécessité de se doter d'une politique d'acquisition, la stratégie

¹²⁶. Cinq organismes documentaires sur les neuf interrogés sur cette question (le service de documentation du *Monde* n'est pas ici comptabilisé).

du « au cas où », dominante dans la profession.

Nous y voyons l'absence de stratégie explicite d'exploration des sources qui ne présente pas le seul inconvénient de générer une forte déperdition d'« énergie » professionnelle, mais, plus fondamentalement, qui enlève au choix des sources en indexation toute possibilité de systématicité et de maîtrise.

II.2.3- Conclusions intermédiaires

En posant comme horizon théorique la notion de « système-archive » proposée par Foucault, nous voulions mettre en valeur :

- d'une part, que les pratiques d'indexation mettent en œuvre des moyens pour réaliser une « transformation discursive », pour établir un niveau proprement documentaire, où un texte peut se concevoir à la fois sous l'angle d'une source (événement) et sous l'angle d'un document (chose). Nous avons sur ce point, par quelques exemples, montré la diversité des mises en discours qui s'observe sous l'adoption d'un même terme d'indexation, ainsi que les traces qu'un document pouvait garder de la source d'où il provient ;
- d'autre part, que cette « transformation discursive » reste, dans la majeure partie des cas, dérégulée : soit parce que les règles explicites qui guident l'exploration des sources en indexation s'appuient sur des langages documentaires et reproduisent ce faisant une « analyse de contenu » dont on a montré précédemment les limites ; soit parce que les règles implicites qui sont mises en œuvre à partir du principe de la confrontation textuelle ne reposent sur aucune stratégie réellement maîtrisée.

Cependant, certains types d'exploration des sources en indexation, notamment ceux qui prennent en compte la spécificité du foyer énonciatif, laissent penser qu'il n'est pas complètement absurde d'envisager les règles d'exploration des sources en indexation dans les termes des formations discursives de Foucault. Une étude précise mériterait d'être menée sur ce point.

Au demeurant, il apparaît que la stratégie d'exploration des sources en indexation telle qu'elle se mène habituellement, si elle reste le plus souvent implicite, n'en produit pas moins des espaces documentaires nécessairement différents d'une pratique documentaire à l'autre. Ce n'est pas cette différence qui nous paraît

problématique ; nous estimons au contraire que la valeur d'un système d'information tient au regard qu'il porte ou qu'il permet de porter sur une production éditoriale. Ce qui nous paraît problématique, c'est que la spécificité de l'espace documentaire créé ne soit pas systématiquement prise en compte, ni au moment de la constitution de la collection, ni au moment de l'attribution des descripteurs. Or, ce n'est qu'en ayant conscience que la sélection des sources aboutit à la création d'un univers documentaire particulier, que l'on peut véritablement poser la problématique de l'usage des mots en indexation : les mots en indexation ont alors pour fonction d'établir un passage entre l'espace documentaire des indexeurs et l'espace d'usages des utilisateurs. En ce sens, tout comme une stratégie d'exploration explicite des sources permet d'avoir une maîtrise de l'espace documentaire que l'on crée, une stratégie d'exposition des documents s'avère nécessaire pour maîtriser le passage d'un espace à l'autre. C'est sur cette problématique que nous nous interrogeons dans le paragraphe suivant.

III- Stratégie d'exposition des documents en indexation

La notion de discours documentaire, entendue comme espace d'organisation spécifique des documents, si elle interroge l'indexation sous l'angle du choix des sources en la contraignant à expliciter ses stratégies d'exploration, permet également de dégager une autre problématique, maintenue elle aussi le plus souvent sous la seule autorité du langage documentaire : comment rendre l'espace documentaire accessible aux utilisateurs ?

Avec la notion de discours documentaire telle que nous essayons de la construire, la problématique de la communication en indexation ne se pose plus, ou plus uniquement, sous l'angle du mot juste ; elle appelle désormais la dimension plus large du discours et de ses conditions d'interprétation. Si, au travers de règles plus ou moins explicites, problématisées et systématiques, l'indexation parvient à transformer des sources hétérogènes en ensembles de documents révélant une certaine cohérence (ne serait-ce qu'« optique » pour reprendre la formulation de Latour), le choix de telles ou telles unités linguistiques pour signaler des liens entre documents peut, au sein du cadre restreint des indexeurs, bénéficier d'une certaine stabilité : mais que devient cette stabilité quand le cercle des locuteurs s'élargit jusqu'aux utilisateurs ? Se dessine la nécessité, pour l'indexation, de disposer d'une stratégie d'exposition des documents qu'elle a créés.

Pour aborder l'aspect de l'indexation sous l'angle d'une stratégie d'exposition des documents, nous procéderons de la façon suivante :

- d'une part, nous disposerons un cadre théorique qui nous permette de problématiser la question des différents « espaces » de locuteurs en jeu dans l'indexation, « espace » des indexeurs et « espace » des utilisateurs ; à ce titre, on empruntera à Kripke son approche des mondes possibles. Cet emprunt, quoique nécessaire, est problématique : issu de la logique, le modèle des mondes possibles, s'il peut être utilisé dans une perspective linguistique, nécessite des reformulations, qui flirtent toujours, dangereusement, avec l'emprunt métaphorique. Nous essayerons d'explicitier ce point et les conditions dans lesquelles nous pourrions utiliser ce concept dans le cadre de notre recherche ;
- d'autre part, compte tenu de la problématique posée, nous présenterons quelques moyens qu'utilise ou que pourrait utiliser l'indexation pour établir des « ponts » entre les différents mondes possibles qu'elle doit traverser.

III.1- La notion de « monde possible » dans le modèle de Kripke

Nous avons tenté de montrer précédemment que la pratique d'indexation exerçait une « mise en discours » des sources au terme de laquelle était créé un espace documentaire. Nous avons utilisé plusieurs formulations pour rendre compte de cette création des documents en indexation¹²⁷. Pour pouvoir poursuivre notre recherche, il nous paraît désormais nécessaire d'essayer de préciser la terminologie employée.

Le recours aux images d'« univers », d'« espace » ou de « monde », pour faire référence à un domaine d'interprétation, est courant¹²⁸. Emprunter le concept logique établi par Kripke et examiner les conditions de son emploi dans une perspective linguistique nous paraissent être de nature à rendre opératoire l'utilisation de ces images, du moins dans le cadre de notre recherche. Le concept de « monde possible » nous permettra notamment de traiter la question du conflit des interprétations entre groupes de locuteurs (indexeurs et utilisateurs) sans poser l'enjeu dans les termes d'objectivité ou de subjectivité, comme il est souvent fait dans la littérature classique sur l'indexation¹²⁹.

III.1.1- Présentation de la notion de « monde possible »

La notion de monde possible proposée par Kripke constitue une « version laïque » de la notion originellement formulée par Leibniz. Si Leibniz oppose le monde réel, créé par Dieu, aux mondes possibles, laissés intentionnellement inactualisés par Dieu, pour Kripke, outre que le monde réel est une sorte de monde possible, le monde possible est toujours une création humaine :

Un monde possible n'est pas un pays lointain qu'on rencontre sur son chemin ou qu'on regarde au télescope. [...] Un monde possible est *donné par les conditions descriptives que nous lui associons*. [...] Les « mondes possibles » sont *stipulés*, ils ne sont pas *découverts* par de puissants télescopes.
Kripke, 1982 [1972], p. 32.

Cette thèse de Kripke s'oppose aussi à celle d'autres logiciens, qui pensent la notion de « mondes possibles » dans les termes d'une « théorie des répliques »¹³⁰. Cette théorie, qui suppose qu'il existe un univers de référence dont on peut imaginer des variantes, se fonde sur la notion d'« identité » entre mondes possibles ; or c'est précisément ce point que met en cause Kripke. En effet, comme nous l'avons

127. Espace de représentation, espace de discours, univers d'interprétation par exemple.

128. On la retrouve par exemple chez Ricœur [1971, p. 175-187] ou chez Rastier [1994].

129. Par exemple Quinn 1994 pour une synthèse.

130. Kripke 1982 [1972], p. 33-34.

précédemment évoqué, sa description de la référence se veut entièrement déconnectée d'une description qualitative, en termes de traits définitoires nécessaires et suffisants :

Même s'il y avait un ensemble purement qualitatif de conditions nécessaires et suffisantes pour être Nixon, la conception que je défends ici n'impliquerait pas qu'il faille trouver ces conditions *avant* d'être en mesure de demander si Nixon aurait pu gagner les élections, ni qu'il faille reformuler la question en termes de telles conditions. Nous pouvons considérer simplement *Nixon* et demander ce qu'il aurait pu lui arriver à *lui* si diverses circonstances avaient été différentes.
Kripke, 1982 [1972], p. 35.

Cet exemple montre que, dans tous les mondes possibles, Nixon reste un être humain : autrement dit, la création de mondes possibles est contrainte. Tout contexte imaginable ne peut fonctionner dans tous les cas comme un monde possible ; il doit, pour constituer un « monde possible », se révéler contrefactuel¹³¹. On doit pouvoir opposer ce qui est possible à ce qui ne l'est pas, c'est-à-dire cerner l'« idée de possibilité qui est en jeu »¹³² dans les mondes possibles créés. Ainsi Kripke propose-t-il de considérer un monde possible comme une « situation contrefactuelle », un « état » et non une entité concrète :

Il suffit de décrire en quoi la « situation contrefactuelle » diffère (de façon pertinente) des faits réels ; on peut concevoir la « situation contrefactuelle » comme un mini-monde ou mini-état, restreint aux aspects du monde qui sont pertinents pour ce qui est en question.
Kripke, 1982 [1972], p. 170.

Selon cette approche des mondes possibles intentionnellement créés, l'« identité » observable entre mondes possibles est une conséquence et non une cause :

Les théoriciens ont souvent dit que nous identifions les objets à travers les mondes possibles grâce à leur ressemblance avec l'objet de départ sous les aspects les plus importants. [...] Au contraire, nous commençons avec les objets, que nous avons et que nous pouvons identifier dans le monde réel. Nous pouvons ensuite nous demander si certaines choses auraient pu être vraies de ces objets.
Kripke, 1982 [1972], p. 41.

Rejetant l'existence *a priori* des mondes possibles et, de ce fait, la nécessité de les décrire « qualitativement », Kripke propose une approche de la notion de mondes possibles qui met en avant deux aspects : la création contrainte des mondes possibles et le caractère résultatif de l'« identité » des objets.

Ces deux aspects sont intimement liés : les « mondes possibles » n'existent pas en soi mais sont toujours créés à partir de la perception d'un objet dont on se demande

131. « Si la proposition *Aristote aimait les chats* est tenue pour factuelle, la proposition *Aristote n'aimait pas les chats* est tenue pour contrefactuelle ». Notons que c'est cette possibilité de construire P et non-P à partir du nom propre *Aristote* qui prouve que le nom propre « n'est pas le sténogramme d'un paquet de prédicats identificateurs ». Milner 1989, p. 331.

132. Kripke 1982 [1972], p. 34.

s'il pourrait être doté de propriétés que l'on ne lui connaît pas dans le monde où l'on se trouve. La démarche de création d'un monde possible ne suppose donc pas une recherche d'identité de propriétés d'un objet ; au contraire, elle stipule que les propriétés d'un objet peuvent changer. Autrement dit, pour Kripke, les critères d'identification d'un objet sont toujours relatives à un monde : ce ne sont pas les mêmes critères que l'on met en œuvre pour distinguer un objet dans le « monde réel » et dans un « monde possible »¹³³. C'est ainsi que, pour Kripke, il ne relève pas d'une théorie des mondes possibles de se demander si tel objet du monde réel (la reine d'Angleterre, dans son exemple¹³⁴) aurait pu être un autre objet (un cygne, dans son exemple) dans un monde possible : ce type de question repose sur l'identification inter-mondes de propriétés semblables d'objets différents et sur l'existence *a priori* des mondes possibles. C'est en ce sens que la construction des mondes possibles est, dans le modèle de Kripke, contrainte. Les contraintes qui pèsent sur la création des mondes possibles sont exprimées en termes d'« essences »¹³⁵ et relèvent de trois ordres¹³⁶ :

- l'origine, par exemple « être humain » pour la reine d'Angleterre : autant on peut créer un monde où la reine d'Angleterre serait une pauvre, autant on ne peut pas créer un monde où elle serait un cygne ;
- la matière, par exemple « être en bois » pour une table en bois : on peut créer un monde où la table en bois dont on parle dans le monde réel est dans un autre endroit que celui que l'on perçoit ; on ne peut, en revanche, imaginer qu'elle puisse fondre (c'est-à-dire imaginer qu'elle soit en glace par exemple) ;
- la forme, par exemple « avoir la forme de table » pour une table : comme précédemment, on ne peut créer un monde possible, au sens où Kripke l'entend, dans lequel une table serait un vase.

Pour autant, ces propriétés essentielles, précise Kripke, ne répondent pas à la question : « quelles propriétés un objet doit-il garder pour ne pas cesser

133. *Ibid.*, p. 37 : « Les propriétés qu'un objet a dans tout monde contrefactuel n'ont rien à voir avec les propriétés dont on se sert pour l'identifier dans le monde réel ».

134. *Ibid.*, p. 100-102.

135. *Ibid.*, p. 41 : « Certaines propriétés d'un objet peuvent lui être essentielles, dans la mesure où il n'aurait pas pu ne pas les avoir. Mais ces propriétés ne servent pas à identifier l'objet dans un autre monde possible, car une telle identification n'est pas requise ».

136. Kripke n'établit pas de cette façon les propriétés essentielles des objets : ses propos sont plus diffus, complexes, nuancés ; pour simplement donner quelques illustrations de ce que semble vouloir dire Kripke sur les propriétés essentielles, on reprend, sans les discuter, la typologie et les commentaires que propose Engel 1985, chap. V.

d'exister ? »¹³⁷, mais plutôt à celle-ci : « quelles propriétés (atemporelles) un objet n'aurait-il pas pu ne pas avoir ? »¹³⁸. La différence entre les deux types de question renvoie à l'existence *a priori* ou pas des mondes possibles : si les mondes possibles sont perçus comme existants *a priori*, ils révèlent une temporalité propre ; pour Kripke, les mondes possibles, n'existant pas *a priori* mais étant construits à partir du « monde réel » (le monde d'où l'on parle, pourrait-on dire), s'introduisent **dans la temporalité**, dans l'« histoire » du monde réel :

D'habitude, lorsque nous nous demandons si, intuitivement, quelque chose aurait pu arriver à un objet donné, nous nous demandons si l'histoire de l'univers aurait pu se dérouler comme elle s'est effectivement déroulée jusqu'à un certain instant, et adopter ensuite un cours différent du cours réel, de sorte que, à partir de là, les vicissitudes de cet objet auraient pu être différentes de celles qu'elles ont été.

Kripke 1982 [1972], p. 104, n. 57.

C'est pourquoi, selon Engel, les propriétés essentielles de Kripke ne constituent pas des propriétés individualisantes, des « essences individuelles ». Elles relèvent plutôt d'essences « sortales » qui « n'ont pas pour fonction d'*individualiser* un individu en lui attribuant une propriété, mais seulement de fixer sa référence en le rapportant à une espèce. On s'assure ainsi que César n'est pas un chien mais un homme »¹³⁹.

La théorie des mondes possibles telle que Kripke la conçoit nous paraît particulièrement pertinente pour notre étude, dans le sens où, d'une part, elle insiste sur le caractère construit et contraint des mondes possibles et où, d'autre part, elle repositionne la question de l'identification des objets entre mondes. Ces deux points sont essentiels pour l'approche de l'indexation que nous proposons.

Avant d'en venir à l'utilisation que nous ferons de la théorie kripkéenne des mondes possibles, il nous faut déterminer dans quelle mesure elle peut être appliquée à des objets linguistiques (des textes, des énoncés, des discours).

III.1.2- Interprétation linguistique de la notion de « monde possible »

Sans souci d'exhaustivité, nous avons relevé dans la littérature que nous avons explorée pour cette recherche, plusieurs types « d'utilisations linguistiques » de la notion de mondes possibles. Ainsi, sont assimilés à des mondes possibles des

¹³⁷. Kripke 1982 [1972], p. 103, note 57.

¹³⁸. *Id.*

¹³⁹. Engel 1985.

textes (la plupart du temps, ce sont les romans¹⁴⁰ qui sont appréhendés en termes de mondes possibles) ou encore des discours¹⁴¹.

Gary-Prieur [1994] ainsi qu'Eco [1985 (1979)] s'attachent particulièrement à justifier de telles utilisations du concept de Kripke : leurs argumentations, si elles conduisent à de semblables conclusions, ne portent pas sur les mêmes éléments.

Gary-Prieur [1994] met en avant que les mondes possibles de Kripke sont « stipulés » et qu'on peut entendre par là qu'ils « n'ont pas d'existence extérieure au discours qui les met en place »¹⁴² ; autrement dit, c'est par le discours que se créent les mondes possibles. Gary-Prieur envisage la notion de monde possible sous l'angle du créateur de ce monde, du côté du locuteur. Elle reste ce faisant très proche des conceptions de Kripke : en effet, comme le souligne l'un de ses commentateurs¹⁴³, les mondes possibles sont, pour Kripke, des « manières de parler », de pures entités linguistiques¹⁴⁴.

Insistant sur un autre aspect de la théorie de Kripke - la notion de situation contrefactuelle -, Eco propose une autre approche des mondes possibles qui ne se situe plus du côté du locuteur mais du côté de l'interlocuteur, pourrait-on dire, du lecteur. Eco considère un monde possible comme un fait d'interprétation : si sa position rejoint celle des auteurs qui abordent le roman en termes de monde possible, elle prend quelque distance avec les propos de Kripke.

Selon Eco, la lecture d'un texte progresse par formulations successives d'hypothèses, de prévisions établies par rapport à ce que le texte a préalablement disposé : un lecteur stipule « un cours d'événements possible ou un état de choses possibles » qui se verront confirmés ou infirmés par la poursuite de la lecture¹⁴⁵. Dans ce cadre, un monde possible est une prévision que le lecteur construit sur la base de ce que dit le texte, qui constitue alors une contrainte sur la création des mondes possibles. Eco insiste sur le fait que l'interprétation d'un texte ne relève pas d'un choix entre alternatives mais d'un choix entre possibilités, définies par le texte lui-même¹⁴⁶. En ce sens, il propose de définir la notion de monde possible comme « un état de choses exprimé par un ensemble de propositions où, pour chaque

¹⁴⁰. Cf. Corblin 1995, p. 198 et suiv. pour des exemples : « Un roman stipule un monde possible », ou encore Eco 1985 [1979], chapitre 8, p. 157-225 : « Structures de mondes », ou Gary-Prieur 1994, p. 26, note 1.

¹⁴¹. Fradin et Marandin [1979] parlent de « reformulation » du concept de Kripke pour étudier les aspects de la rigidité comme effets de référence dans le discours.

¹⁴². Gary-Prieur 1994, p. 21.

¹⁴³. Engel 1985, chapitre IV.

¹⁴⁴. *Id.* : « Kripke semble se ranger dans le camp de ceux qui considèrent le langage en termes de mondes comme heuristique et réduisent ceux-ci à de pures entités linguistiques ou des manières de parler ».

¹⁴⁵. Eco 1985 [1979], p. 146.

¹⁴⁶. *Ibid.*, p. 160.

proposition, soit p soit non- p »¹⁴⁷. Sur ce point, Eco rejoint en partie Gary-Prieur, bien que leur cadre d'approche soit radicalement différent (sémiotique des textes narratifs pour Eco, grammaire du nom propre pour Gary-Prieur) : c'est à l'intérieur d'un énoncé que peuvent intervenir plusieurs mondes possibles¹⁴⁸.

Compte tenu des caractéristiques que Kripke attribue à la notion de monde possible (construction contrainte), il semble légitime de pouvoir associer la notion de monde possible à la notion linguistique de construction référentielle : un monde possible correspond alors à la construction de la référence en discours. C'est sur cette base que nous nous appuierons pour élaborer notre propre approche de la notion de monde possible.

III.1.3- Enjeu de la notion de « monde possible » en indexation

Cet essai d'interprétation des enjeux de l'indexation dans le cadre de la théorie de Kripke a pour objectif de repositionner la problématique des mots en indexation au regard de la question de l'identité de l'objet telle que la pose Kripke. Il s'agit de montrer que les descripteurs ne peuvent avoir pour rôle d'identifier le « même » objet dans tous les mondes possibles, mais qu'ils ont plutôt pour fonction d'établir le passage d'un objet d'un monde possible à l'autre, du monde des indexeurs à celui des utilisateurs.

Sur quelles bases peut-on assimiler le cercle des indexeurs à un monde possible et celui des utilisateurs à un autre monde possible, ou plutôt à un ensemble de mondes possibles ?

On ne discutera pas les approches linguistiques précédemment présentées : on admettra que la construction d'un monde possible repose sur une formulation linguistique. Mais nous focaliserons, quant à nous, notre attention sur le rôle de l'objet dans la construction du monde possible.

A- Univers des documents et création d'un monde possible

Si l'on considère que l'ensemble des productions éditoriales constitue, pour les indexeurs, le « monde réel », on pourrait penser que l'univers des documents n'est qu'une portion, qu'une partie, de ce monde réel, mais qu'il n'en propose pas, à proprement parler, un autre « état », pour reprendre les termes de Kripke. Or, nous avons tenté de montrer que la sélection des sources en indexation peut se

¹⁴⁷. *Ibid.*, p. 165.

¹⁴⁸. Gary-Prieur 1994, p. 22.

comprendre comme une décontextualisation qui pouvait permettre à une source de se voir attribuée d'autres propriétés, notamment celles de nouveaux usages. La création de l'univers des documents en indexation nous paraît en cela pouvoir être assimilée à la création d'un monde possible : construit à partir des objets du « monde réel » (les sources), le monde possible des documents attribue de nouvelles propriétés à ces objets. L'attribution de ces nouvelles propriétés ne se fait pas dans le cadre d'un discours compris au sens linguistique du terme comme production d'énoncés. La notion de discours documentaire que nous avons proposée se rattache plutôt à la notion de « transformation discursive » de Foucault, qui modifie la situation d'énonciation plutôt que l'énoncé lui-même. En cela, notre conception de la notion de monde possible peut sembler s'éloigner à la fois de celle de Kripke et des reformulations linguistiques dont elle a pu faire l'objet. Néanmoins, il nous semble que nous gardons de Kripke l'idée que la construction d'un monde possible s'établit sur la base d'un objet du monde réel dont on se demande quelles autres propriétés (ici d'usage) il va pouvoir se voir attribué. Par ailleurs, si l'indexation ne produit pas de discours à proprement parler, les descripteurs peuvent être considérés comme l'ancrage linguistique d'un monde possible.

B- Les mondes possibles des utilisateurs

Là encore, notre utilisation de la notion kripkéenne de monde possible repose essentiellement sur la spécificité de la construction d'un monde contraint par un objet du monde réel. Dans le cas des utilisateurs, on considérera que le monde réel est le monde d'où ils parlent, dans lequel ils se situent. Une situation de recherche documentaire peut se comprendre comme la construction d'un monde possible si l'on considère qu'une requête est établie à partir d'un objet connu (ici un objet de discours que l'on peut nommer) dont on cherche des propriétés inconnues du monde où l'on est¹⁴⁹. Une requête comme, par exemple, « les expéditions dans le désert », revient à imaginer qu'il existe un monde où cet objet de discours est doté de propriétés qui ne sont pas celles du monde où l'on se place. En ce sens, l'utilisateur espère, d'une certaine façon, trouver dans un système d'information ou une bibliothèque, un « état » du monde qui n'est pas exactement celui qu'il connaît.

¹⁴⁹. Voir sur ce point Le Guern 1991b, p. 71 : « Le contenu informatif d'un document peut être analysé comme constitué de mises en relations de propriétés avec des entités, objets concrets ou non : un objet de pensée, une abstraction, est un objet logique tout autant que les choses du monde. On peut représenter la démarche de l'utilisateur d'un système d'information comme la détermination de l'objet à propos duquel il désire trouver des renseignements, puis la recherche des propriétés attribuées dans les documents du corpus à cet objet, ce qui passe par le repérage des documents qui font mention de cet objet ».

C- Problématique des relations entre mondes possibles en indexation

Si l'on accepte de considérer que le cercle des indexeurs constitue un monde possible et que le cercle des utilisateurs suppose un ensemble d'autres mondes possibles, la question qui se pose à l'indexation est celle de la compatibilité entre ces mondes ou encore le passage d'un monde possible à l'autre.

Sur ce point, Kripke souligne que le passage d'un monde à l'autre ne peut se fonder sur l'identification des propriétés d'un objet. Dans son modèle, deux mondes différents ne peuvent être liés sur la base d'une recherche d'identité ; un monde possible est toujours créé à partir d'un monde posé comme réel. Or, dans l'indexation, on aurait deux mondes construits chacun de leur côté sur la base d'objets différents, ce qui rend *a priori* la relation impossible. Il faut donc supposer que l'un des deux mondes, celui des indexeurs ou celui des utilisateurs, contraint l'autre. Se dégage alors l'alternative suivante : soit c'est le monde possible d'un utilisateur qui contraint l'univers documentaire ; soit c'est l'univers documentaire qui contraint la construction de l'univers référentiel des utilisateurs. Si elles sont symétriquement opposables, ces deux solutions ne mettent pas en cause les mêmes enjeux.

Nous examinerons successivement les deux branches de cette alternative.

C.1- La première - les mondes possibles des utilisateurs contraignent la création de l'univers de référence documentaire - constitue, nous semble-t-il, le modèle dominant en indexation, pour peu que l'on reformule l'approche classique de l'indexation dans le cadre de la théorie de Kripke. Il consiste à intégrer, au sein d'un seul monde possible - celui des indexeurs - le maximum d'autres mondes possibles, c'est-à-dire, le plus souvent, à prévoir, par les descripteurs, le maximum de possibilités d'interrogation. Cette solution ne prend pas en compte le fait que l'univers des documents est construit et à ce titre doublement contraint : par les sources elles-mêmes et par les choix que sont conduits à faire les indexeurs. À moins de confondre bibliothèque et librairie, le choix d'acquisition en bibliothèque ne se réduit jamais aux propositions d'achats des lecteurs. Pour peu que l'on pose la question du document en indexation, il apparaît que les mondes possibles des utilisateurs (les « besoins d'information ») ne peuvent être pris en compte au niveau de l'univers des indexeurs lui-même.

C.2- S'il y a nécessité de rapport contraint entre monde des indexeurs et monde des utilisateurs, la contrainte ne peut s'exercer que dans un seul sens : c'est le monde des indexeurs qui contraint celui des utilisateurs. Dès lors, l'enjeu du descripteur ne se pose plus de la même façon, et d'autres dimensions, que celle des mots, doivent

être mises en œuvre. L'opération d'indexation doit disposer de quoi guider l'interprétation, de quoi contraindre la lecture. On dira qu'à ce titre elle met en place une « stratégie d'exposition » des documents. Sur ce point, on rejoint la conception de monde possible proposée par Eco : l'utilisateur se trouve dans la position d'un lecteur qui progresse dans sa recherche d'informations par formulations successives d'hypothèses qui lui sont proposées par la collection documentaire, alors perçue comme « monde réel ». Dans ce cadre, le descripteur n'a plus pour rôle de « prévenir » la formulation linguistique des requêtes d'un utilisateur. L'enjeu ne consiste plus à anticiper la variation linguistique mais à l'exploiter, au contraire.

En ce sens, une approche des fondements théoriques de l'indexation peut s'inscrire dans un programme de recherche tel que le conçoit Dubois :

Prendre la mesure des variations en langue et langages comme une donnée de fait, sinon comme l'essence même des phénomènes de langue [...] ne pas traiter la diversité des interprétations en termes de décalage par rapport à une vérité sémantique du texte mais par rapport à une norme d'interprétation historiquement et socialement située .
Dubois 1995, p. 93.

L'enjeu de l'indexation revient donc à exposer comme telle la « norme d'interprétation » des documents qu'elle propose, norme dans laquelle les utilisateurs pourront alors se situer.

Le recours au modèle des mondes possibles établi par Kripke nous a paru indispensable pour poser les enjeux de l'indexation relatifs à la communication entre indexeurs et utilisateurs. Il ne permet pas, bien sûr, d'envisager l'intégralité des aspects de cette relation. Néanmoins, par le modèle de Kripke, le rapport entre indexeurs et utilisateurs peut se percevoir sous l'angle de la problématique précise de la compatibilité entre univers de référence : Kripke établit une notion de monde possible qui exige un ancrage dans un monde réel. Cette conception nous paraît précieuse notamment par la reformulation de la question de l'identité d'un objet entre mondes qui la sous-tend : si l'on peut parler d'un même objet, d'un monde à l'autre, c'est parce que nous-mêmes l'aurons ainsi décidé, dit en substance Kripke. Cette proposition souligne la nécessité en indexation de forcer les différents mondes à se rencontrer, sans pouvoir s'en remettre aux choses telles qu'elles se donnent.

III.2- Éléments pour une stratégie d'exposition des documents en indexation

En proposant de concevoir l'indexation à travers la notion de stratégie d'exposition des documents, nous essayons de déplacer la problématique classique de l'indexation. Il ne s'agit plus de poser l'existence *a priori* d'objets semblables dans différents mondes (celui des auteurs, des indexeurs, des utilisateurs) et de trouver un mode de désignation commun à tous ces mondes ; il s'agit, au contraire, de partir des objets tels qu'ils se donnent, dans leur singularité, et de définir les moyens qui pourraient permettre de contraindre le regard porté sur ces objets.

Nous entendons, par stratégie d'exposition, l'établissement d'un cadre au sein duquel les utilisateurs peuvent construire leur parcours interprétatif : le cadre d'interprétation fonctionne comme une contrainte interprétative en indiquant dans quel « domaine » ou sous quel « angle » un ensemble de documents peut être perçu.

Quels sont les moyens que l'indexation peut mettre en œuvre pour disposer un cadre qui contraigne l'interprétation des documents ?

On trouve, dans la pratique documentaire classique, les prémices d'une stratégie d'exposition des documents qui recourt aux langages documentaires, et, plus particulièrement, aux langages classificatoires (III.2.1). Ce type de stratégie conduit à des limites que nous avons déjà signalées¹⁵⁰. Un autre type de stratégie pourrait être envisagé qui s'inspire de la démarche adoptée par les vulgarisateurs scientifiques (III.2.2).

III.2.1- Le discours classificatoire

Les stratégies d'exposition des documents s'appuient, dans les pratiques courantes, sur l'utilisation de classifications documentaires. Après avoir montré en quoi celles-ci pouvaient se comprendre comme des « stratégies », nous examinerons deux aspects du discours classificatoire : les classifications hiérarchiques et les classifications à facettes.

A- Les classifications comme « stratégies d'énonciation de l'offre documentaire »¹⁵¹

L'utilisation des classifications comme modes d'exposition des documents est particulièrement visible dans les bibliothèques qui ont opté pour la mise en accès

¹⁵⁰. Voir § II.2.2 dans ce chapitre.

¹⁵¹. L'expression est d'É. Véron 1990.

direct des documents¹⁵². Comme précédemment la notion de politique d'acquisition, le libre-accès des documents constitue une préoccupation récente en France¹⁵³, qui fait, depuis ces quinze dernières années, l'objet de débats : faut-il maintenir la classification Dewey pour présenter les documents au public ou faut-il créer de nouveaux modes d'exposition¹⁵⁴ ?

L'introduction du libre-accès en bibliothèque a élargi la problématique de l'exposition à d'autres dimensions que celle du seul langage documentaire. En effet, il est très vite apparu que la logique de la classification Dewey, « unidimensionnelle », ne pouvait apporter toutes les réponses nécessaires à un aménagement de l'espace, lui « tridimensionnel » : dès lors « la classification elle-même ne comportant aucune règle d'étalement spatial, un même fonds peut être spatialisé d'une multitude de manières différentes »¹⁵⁵.

En examinant la mise en espace des collections dans quatre bibliothèques publiques, Éliséo Véron [1990] a pu dégager, à partir de l'utilisation de la classification Dewey, des « discours stratégiques » différents, le plus souvent implicites. La classification Dewey apparaît ici, comme précédemment pour mener une politique d'acquisition, comme un instrument au service d'une stratégie qui ne se dit pas. Mais, ici encore, le moyen finit par prendre plus d'importance que la « fin ».

Au cours de son étude, Véron relève que la diversité des mises en espace n'influe en rien sur les stratégies d'appropriation de la bibliothèque par les lecteurs, comme si la classification constituait déjà, en elle-même, une mise en espace, dont la « qualité » intrinsèque influe peu sur le caractère opératoire :

Nous sommes tentés de conclure à ce propos que le rapport des usagers à la classification implique tout simplement la nécessité d'un système de repérage permettant la constitution d'une stratégie. La classification est une garantie de l'existence d'une convention particulière, de l'absence d'arbitraire. Autrement dit, une classification est indispensable, ne serait-ce que comme élément contre lequel organiser une stratégie, mais cette classification ne nécessite pas des perfectionnements particuliers, et l'on peut soupçonner que n'importe quelle classification, pourvu qu'elle soit stable et régulière, fasse l'affaire.
Véron, 1990, p. 81.

¹⁵². Les principes sont les mêmes, la visibilité en moins, dans les systèmes d'information.

¹⁵³. Les débuts du libre-accès se situent après la Seconde Guerre mondiale, un développement plus massif se note dans les années 1970 ; toutes les bibliothèques ne proposent pas, à l'heure actuelle, de collections en libre-accès.

¹⁵⁴. Certaines bibliothèques proposent des modes d'exposition des documents par « centres d'intérêt », qui réorganisent les classes de la Dewey : il s'agit là d'un choix « orienté utilisateurs » (« Le centre d'intérêt est un espace logique et matériel dont le véritable centre est le lecteur », Véron 1990, p. 88). Une structuration par centres d'intérêt peut fournir des catégorisations de type : « pays, paysages, voyages » ; « le monde des spectacles » ; « l'art et les artistes » ; « vécu » ; « le temps libre et vos loisirs », etc. (cité in Véron 1990, p. 85).

¹⁵⁵. *Ibid.*, p. 11.

Il nous semble que ce propos de Véron souligne le fait qu'une classification fonctionne déjà, en soi, comme une stratégie d'exposition. En effet, ce que ne considère pas toujours Véron dans son étude, c'est que la classification, si elle est utilisée par les professionnels des bibliothèques dans le cadre de stratégie plus ou moins explicite, est avant tout construite par les professionnels eux-mêmes et constitue à ce titre un premier niveau de « discours », le « discours classificatoire ». Dès lors, les observations que Véron formule sur l'efficacité de la seule classification peuvent être vues, moins comme un échec des différentes tentatives de mises en espace des collections, que comme un signe de la nécessité absolue de disposer d'un cadre dans lequel les utilisateurs puissent construire leur parcours, leur système d'interprétation.

Véron semble dire que tous les cadres se valent, même si sa préférence va aux cadres les plus « décalés »¹⁵⁶. La question nous semble plus ouverte, ne serait-ce que parce que les différents types de classifications eux-mêmes n'établissent pas exactement le même type de « discours », de stratégie d'exposition.

Dans les deux paragraphes suivants, nous essayerons de montrer comment se manifeste cette diversité du « discours classificatoire ».

B- Les classifications hiérarchiques

On reprend de Bourion et Malrieu [1994] leur approche du plan de classification en termes de « discours classificatoire ». Elles définissent le plan de classification comme un « texte techno-scientifique utilisant une terminologie consensuelle, celle des langues de spécialité qui relèvent de la discipline traitée »¹⁵⁷. Elles s'interrogent sur la façon dont une classification gère les objets particuliers que sont les textes et posent l'hypothèse qu'un plan de classement, s'il s'apparente au modèle canonique des classifications (logique de type inclusion de classes), repose en fait sur une sémantique de relations tout à fait particulière, qui spécifie un type de discours particulier, le discours classificatoire :

[Un plan de classification] se présente comme une arborescence qui évoque une structure logique canonique de type inclusion de classes, mais la nature ontologique des objets classés (« objets sémiotiques » et non « objets du monde physique ») va

¹⁵⁶ *Ibid.*, p. 86 : « Les résultats de cette recherche nous ont amenés à une conclusion : le système de classification lui-même est moins important que ne le laissent supposer les discussions passionnées autour de la Dewey. [...] L'enjeu véritable de l'opposition entre les partisans de la classification Dewey (plus ou moins "adaptée") et les partisans de la philosophie des "centres d'intérêt", est celui de choisir entre une classification marquée par une conception du monde et des savoirs qui datent du XIX^e siècle, et une classification qui sera inévitablement marquée par une autre idéologie plus actuelle. À cet égard, on peut se demander s'il n'est pas préférable de faire appel à une classification en décalage avec le monde contemporain (comme la Dewey) plutôt qu'à un système qui ne fait que reproduire (cette fois dans l'espace des bibliothèques municipales) la "grille" consacrée autour de nous ».

¹⁵⁷ Bourion et Malrieu 1994, p. 84. Nous ne rentrons pas ici dans le détail de leur étude.

de pair avec des dimensions de description obligatoirement hétérogènes, en contradiction avec la structure arborescente taxinomique.
Bourion et Malrieu, 1994, p. 84.

L'étude que mènent Bourion et Malrieu [1994] du plan de classement utilisé par le centre de documentation du CNRS dans le domaine de la psychologie montre que, sous l'adoption explicite d'un cadre canonique et objectif, se déploie un discours classificatoire spécifique aux indexeurs :

- les indexeurs créent des classes ou des sous-classes qui ne se réfèrent pas à des champs de spécialités reconnus par les experts d'un domaine mais qui permettent de ranger des documents qui se situent aux confluent de plusieurs problématiques ; Bourion et Malrieu dégagent sur ce point une sémantique de relations spécifique aux indexeurs, qui ne doit rien à la logique d'inclusion de classes¹⁵⁸ ;
- s'ils modifient la modélisation scientifique d'un domaine, les indexeurs ne se sentent pas toujours tenus d'introduire dans un plan de classement de nouvelles dénominations de classe ; il se passe souvent plusieurs années avant qu'une nouvelle dénomination scientifique soit intégrée dans un plan de classement¹⁵⁹.

Un certain nombre de modifications et de choix sont donc effectués par les indexeurs sur la base d'un référentiel scientifique : c'est ce qui constitue le « discours classificatoire ». Ces changements imposés, insistent Bourion et Malrieu, par la spécificité sémiotique des objets que manipulent les documentalistes, restent cependant peu explicités et problématisés, notamment le rôle implicite attribué au graphisme dans un plan de classement ne fait l'objet d'aucune verbalisation¹⁶⁰. Nous dirons qu'une classification constitue en cela une stratégie d'exposition implicite des documents, stratégie pas toujours maîtrisée comme telle, empruntant ses moyens non seulement à la langue mais aussi à d'autres systèmes sémiotiques (comme le graphisme).

La particularité du discours classificatoire, un discours « non dit », n'est pas sans incidence sur l'interprétation que peuvent faire les utilisateurs des dénominations des classes d'un plan de classement : les termes spécialisés utilisés dans la classification perdent peu à peu tout lien avec le domaine d'où ils viennent sans que l'utilisateur ne dispose de nouveaux cadres explicites d'interprétation¹⁶¹. Faut-il

¹⁵⁸. *Ibid.*, p. 93-104.

¹⁵⁹. Bourion et Malrieu donnent l'exemple du terme « victimologie », repéré dans la littérature dès 1981, mais définitivement introduit dans le plan de classement en 1983. *Ibid.*, p. 87.

¹⁶⁰. *Ibid.*, p. 116.

¹⁶¹. *Ibid.*, p. 117.

alors « intégrer dans les graphes conceptuels les contraintes liées au genre et au domaine », comme le proposent les deux auteurs à la toute fin de leur étude¹⁶² ? Il ne nous semble pas que ce soit au niveau du langage documentaire que la contextualisation des termes soit la plus efficace. En effet, il nous semble que l'on rencontre alors le problème initial soulevé par Bourion et Malrieu concernant la spécificité sémiotique des objets de l'indexation : on ne peut classer un document comme on classe des objets, des mots, des concepts. Si les termes scientifiques peuvent être classés dans un domaine (une terminologie), parce que ce sont finalement, dans ce cas, des concepts scientifiques (des « choses » pourrait-on dire) qui sont classés, il ne semble pas que des documents, compte tenu de leur spécificité « textuelle », puissent être, au même titre que des termes, classés dans un domaine. Une autre approche devrait pouvoir être dégagée pour tenir compte de la spécificité des textes ; nous en proposons des aspects en II.2.2.

Dans le discours classificatoire fondé sur un langage documentaire, on se heurte à nouveau à deux types de problèmes que l'on a pu régulièrement identifier au cours de cette étude :

- d'une part, le rapprochement explicite de textes se fait *a posteriori* lorsque les textes ont déjà été sélectionnés et distribués en grandes disciplines : stratégie d'exploration des sources et stratégie d'exposition des documents ne sont pas considérées dans leur spécificité ;
- d'autre part, et compte tenu du fait que l'on ne dispose d'aucun *a priori* non formel pour dicter le rapprochement des documents, on en vient à être contraint de réduire un texte non seulement à un mot (indice de classification) mais aussi à un domaine.

Si les classifications hiérarchiques peuvent, en indexation, indiquer la présence d'un discours classificatoire qui tente d'établir une « communication » sur la base d'un référentiel scientifique, elles ne peuvent à proprement parler constituer une stratégie d'exposition des documents qui permette une libre circulation entre textes : la contrainte du langage documentaire comme l'approche en termes de contenu y restent trop fortes.

Une variante du discours classificatoire, articulée sur la notion de « facettes », pourrait-elle être plus efficacement utilisée dans une stratégie d'exposition des documents ?

¹⁶². *Id.*

C- Les classifications à facettes

La notion de facette, élaborée par Ranganathan¹⁶³ pour optimiser le classement des ouvrages en bibliothèque¹⁶⁴, peut s'approcher en termes de « point de vue » sur un objet.

La spécificité des langages documentaires à facettes tient, d'une part, à ce qu'un document y est systématiquement envisagé sous plusieurs angles, et, d'autre part, à ce que les facettes relèvent d'un type de propriétés particulières d'un objet : les facettes s'apparentent en effet aux propriétés « sortales » et non individuelles d'un objet¹⁶⁵. La notion de facette vise à introduire, dans les langages documentaires, une autre approche que celle du seul contenu : c'est à partir de propriétés extralinguistiques qu'un objet est classé.

Sans rentrer dans le détail des facettes proposées par Ranganathan¹⁶⁶, on peut donner un exemple¹⁶⁷ de ce que tend à proposer une approche par facettes :

- par la facette Personnalité, le terme « rose » pourra être appréhendé sous l'angle de « l'espèce végétale » ;
- par la facette Matière, le terme « instrument » pourra être perçu sous l'angle de la « musique » ;
- par la facette Énergie, le terme « labourage » sera perçu sous le point de vue « agriculture » ;
- les facettes Espace et Temps reprennent les coordonnées spatio-temporelles indiquées dans un document.

On remarque que le principe de l'analyse par facettes rejoint celui de l'inscription d'un terme dans un domaine, mais la différence tient à ce que plusieurs domaines peuvent être convoqués pour le classement d'un même document¹⁶⁸, même si l'analyse par facettes suppose toujours un rattachement préalable à une classe

¹⁶³. Mais dont, comme le suggère Michel Le Guern, on peut identifier les prémices chez Bernard Lamy. Voir Mustafa-Elhadi [1989, p. 180 et suiv.] qui étudie cette filiation historique.

¹⁶⁴. Directeur de la bibliothèque de l'Université de Delhi puis organisateur du réseau des bibliothèques en Inde, Ranganathan crée en 1933 le premier langage documentaire à facettes, la *Colon Classification*, voir Ranganathan 1976.

¹⁶⁵. Ranganathan reprend d'Aristote cinq catégories susceptibles d'appréhender un même objet sous différents points de vue ; ces catégories ont été traduites par les termes français suivants : Personnalité, Matière, Énergie, Espace, Temps ; [Ranganathan 1976].

¹⁶⁶. La littérature sur le sujet est abondante. On peut se reporter par exemple à Vickery 1963 qui propose les plus importantes reformulations des principes de Ranganathan, à de Grolier 1962, 1970 et 1988 pour une mise en perspective, à Salvan 1972 pour une approche comparative, à Mustafa-Elhadi 1989 pour une approche historique et pour une reformulation des concepts en terminologie.

¹⁶⁷. Repris de Salvan 1972, p. 32-33.

¹⁶⁸. Le classement d'un document s'effectue sur la base de l'analyse de son titre, plus ou moins enrichi par la lecture de certaines parties du document et plus ou moins réécrit, Ranganathan 1976.

principale¹⁶⁹.

Ainsi l'indice suivant construit à partir de la *Colon Classification* : « NA 561, J 37, 67 : 8 » devra-t-il se lire « maquette de tour d'un château Tudor » avec, pour cadre d'interprétation, la classe « beaux-arts ». Chaque élément de l'indice est pourvu d'une signification autonome qui doit permettre une combinaison de sujets. Ainsi « NA 561 », qui se lit « Angleterre, Architecture », devra-t-il permettre de retrouver le document sur la « maquette de tour d'un château Tudor » ; de même de la suite « NA 561, J 37, 6 », qui se lit « Toit, château, Tudor (période) ». Autrement dit, l'indice de base génère plusieurs indices (par « découpage » successif) qui constituent autant d'accès, pour un utilisateur, au même document, selon la facette envisagée (matériau : toit ; temps : Tudor, etc.).

Le principe des facettes introduit par Ranganathan revient donc à restituer, pour chacun des mots retenus dans la définition du sujet d'un document, un contexte d'interprétation de type référentiel, mais là encore ce sont plus les interprétations d'un document que le document lui-même qui sont classées, sans que le point de vue de l'interprétation (la stratégie de lecture adoptée) ne soit, lui-même, spécifié.

Reste qu'en matière de stratégie d'exposition, la notion de facette représente un niveau supplémentaire de transparence par rapport aux précédentes classifications hiérarchiques : elle explicite le point de vue sous lequel il convient de lire un terme, sans laisser à l'utilisateur le soin de se perdre dans le parcours d'une structure arborescente qui répond à la seule logique des indexeurs.

Sans doute une stratégie d'exposition des documents qui reposerait sur le seul usage d'un langage documentaire aurait-elle intérêt à utiliser le principe des facettes, ne serait-ce que parce que l'inscription des termes d'indexation dans un domaine d'interprétation y est systématiquement explicitée. Une étude précise reste à faire sur ce point, qui devrait cependant tenir compte du contexte qui, à l'époque, a présidé à l'instauration du principe des facettes, de façon à écarter ce qui désormais peut se formuler autrement tout en maintenant, parallèlement, les principes de Ranganathan qui restent valides. En effet, à la lumière de ce que nous connaissons désormais de la construction des unités terminologiques, il semble que l'on peut reformuler fructueusement l'approche de Ranganathan, souvent bien complexe dans son système de notation et trop *ad hoc* dans la définition des facettes.

¹⁶⁹. Pour exemple, on ne donne ici que le premier niveau des classes principales de la *Colon Classification* : « Généralités ; Sciences mathématiques ; Sciences physiques ; Expérience spirituelle, mysticisme ; Humanités et sciences sociales ».

En effet, comme le rappelle Vickery :

Les raisons pour lesquelles l'ancienne forme de classification énumérative, symbolisée par l'arbre de la science, est actuellement périmée sont assez claires. Les sujets hautement spécifiques qui font aujourd'hui l'objet du catalogage sont des sujets *composés* ; ils ne peuvent être correctement exprimés que par des vedettes-matières combinant deux termes ou plus de deux termes. Chaque terme peut être utilisé dans une grande variété de combinaisons. Aussi est-il nécessaire d'assurer à cette formation une flexibilité totale. C'est la seule manière d'assurer une référence spécifique aux innombrables sujets particuliers. D'autre part, les utilisateurs exigent également de pouvoir, dans une série donnée, disposer de possibilités exhaustives pour conduire une recherche générique ; ils veulent pouvoir repérer un document sur un sujet complexe spécifique non seulement lorsque ce sujet particulier fait précisément l'objet de leur recherche mais aussi lorsque cette recherche porte sur n'importe quel terme collectif incluant l'un des termes de la composition. Ceci implique non seulement que les termes se prêtent à des combinaisons illimitées mais aussi que les relations génériques (ou relations de classes) soient incluses dans la structure du système. La classification « à facettes » est un moyen d'atteindre ces résultats.
Vickery, 1963, p. 4.

Par cette longue citation de Vickery, nous voulions indiquer que si, dans les années 30 comme dans les années 60, la classification à facettes pouvait apparaître comme l'un des rares moyens (voire le seul) permettant d'exprimer des sujets « composés », les connaissances actuelles issues de la terminologie peuvent nous permettre d'apporter un autre type de réponse au problème qu'exprime ici Vickery. En effet, la terminologie a pu se livrer à l'étude d'emboîtement de termes (on dit, par exemple, que le terme « base de données » est inclus dans le terme « système de gestion de bases de données ») ainsi qu'à l'étude du rôle classificatoire des « têtes » des unités terminologiques. Sans doute pourra-t-on trouver là des pistes qui permettent de repenser la définition initiale des systèmes à facettes. Nous creuserons certaines d'entre elles dans le chapitre V, mais pas sans avoir préalablement observé le rôle des terminologies dans le discours vulgarisant.

III.2.2- Le discours vulgarisant

Précédemment¹⁷⁰, nous avons émis l'hypothèse que la vulgarisation scientifique et l'indexation présentaient une problématique similaire : celle du passage d'un seul monde possible, où la référence des termes est fixe, à une pluralité de mondes possibles, où la référence des termes est variable. C'est à ce titre que nous pensons que l'examen des principales stratégies discursives à l'œuvre dans le discours de vulgarisation scientifique peut nous aider à formaliser ce que pourrait être une stratégie d'exposition en indexation.

¹⁷⁰. Voir § I.3 dans ce chapitre.

Cette comparaison ne vaut que si les « corpus » sont en vulgarisation comme en indexation établis sur le même principe. En vulgarisation, l'analyste établit un corpus sur la base de la recherche des « formations discursives »¹⁷¹. En indexation, nous avons vu qu'il ne pouvait s'agir que d'un idéal théorique dont on voit encore mal, pour le moment, une application généralisée. Reste que nous prendrons comme hypothèse, dans ce paragraphe, que les documents sont ici regroupés dans un « système-archive ». Tenir cette hypothèse permet de maintenir distinctes stratégie d'exploration des sources et stratégie d'exposition des documents.

Cette hypothèse posée, on peut examiner comment le discours vulgarisant procède, de façon à dégager quelques éléments pour élaborer une stratégie d'exposition propre à l'indexation.

A- Stratégie d'exposition en vulgarisation scientifique

Comme s'attache à le montrer Jacobi, la diffusion de connaissance ne se fait pas sans le recours à une stratégie d'exposition¹⁷². La science des chercheurs ne se laisse pas naturellement « exposer » ; il faut qu'elle soit « mise en scène », ce qui suppose un traitement particulier des terminologies¹⁷³.

En effet, comme nous l'avons précédemment relevé, le discours de vulgarisation scientifique ne peut, au risque de se nier, ignorer les termes des terminologies spécialisées. Or ceux-ci n'existent et ne font sens que dans le cadre d'un cercle restreint de locuteurs : sortis du monde où ils sont « possibles », ils perdent une grande partie de leur efficacité¹⁷⁴. Le problème consiste donc à maintenir, dans le cadre d'un discours de vulgarisation à large spectre, la présence des termes sans pour autant laisser se dissoudre leur référence.

L'essentiel du travail de vulgarisation consiste alors en une série de reformulations engagées autour des termes, considérés comme des « termes-pivots » : « les mots savants représentent en quelque sorte les traceurs d'une activité de reformulation en train de se faire »¹⁷⁵.

Deux principaux types de procédé linguistique sont utilisés - les procédés définitionnels et les procédés désignationnels - qui opèrent, tous les deux sur la

171. *Supra*, § I.4.

172. Jacobi 1987.

173. Mortureux 1988, p. 124.

174. Jacobi 1993, p. 74 : « Si personne ne discute l'efficacité des termes spécialisés au sein de petites communautés sociolinguistiques d'experts qui les mobilisent, on sait que, dans une perspective de socio-diffusion, à destination d'un plus grand nombre d'interlocuteurs, non-spécialistes ou novices, ces terminologies cessent d'apparaître comme un excellent vecteur communicationnel pour se muer en obstacle ».

175. Jacobi 1987, p. 64.

référence :

- le paradigme définitionnel consiste à établir une « co-occurrence dans un champ discursif donné de plusieurs définitions, paraphrases, gloses, non identiques (formellement, et parfois même sémantiquement) »¹⁷⁶ ;
- le paradigme désignationnel s'apparente au « déploiement d'un ensemble de désignations qui sont approximativement co-référentielles »¹⁷⁷.

Il nous semble que c'est particulièrement le principe de la constitution de paradigmes désignationnels qui pourrait être exploité en indexation.

En vulgarisation, le paradigme est organisé par le discours qui dispose, autour d'un terme-pivot, d'une série de désignations qui, en contexte, constituent des équivalents approximatifs du terme scientifique visé. Ce que montre la vulgarisation, c'est qu'un mot ne remplace pas un autre mot, mais qu'au contraire les saisies successives que permettent des désignations multiples sont les seules façons de s'approprier la « connaissance » ; comme le note Mortureux, « les locuteurs peuvent jouer du vocabulaire pour désigner une seule et même réalité [...] sans que cette diversité de vocables semble gêner l'identification du référent »¹⁷⁸.

Sur un plan linguistique, un paradigme désignationnel est constitué d'un ensemble de syntagmes nominaux fonctionnant, dans un corpus donné, en co-référence avec un syntagme nominal pourvu, lui, dans les discours spécialisés, d'une définition explicite¹⁷⁹. Il y a donc nécessairement présence, dans le discours de vulgarisation, de textes de nature différente, en majorité des textes de nature scientifique. Les « appels » d'un texte à l'autre, parfois explicites (mention de sources ou citation d'auteurs), le plus souvent implicites, se font sur la base des termes spécialisés : en cela, le discours de vulgarisation procède par « focalisations » sur des termes-pivots.

Il n'est pas le lieu de décrire ici tout le « répertoire métalinguistique »¹⁸⁰ dans lequel peut piocher le vulgarisateur. Il nous importe surtout d'illustrer, à ce stade, la possibilité que des procédés linguistiques de reformulation, eux-mêmes contraints par la langue¹⁸¹, contraignent la lecture et orientent l'interprétation, celles-ci étant

176. Mortureux 1993, p. 3.

177. *Id.*

178. Mortureux et Petit 1989, p. 47.

179. *Ibid.*, p. 44.

180. Jacobi, 1993, p. 69-83.

181. Cf. Mortureux 1993, p. 6 : « Il s'agit là d'une liberté **stratégique** au moins **partiellement déterminée par la structure du lexique** » (c'est nous qui soulignons). Mortureux montre comment l'activité professionnelle de reformulation linguistique faite par le

guidées par les unités lexicales particulières que sont les « termes ».

L'exemple du discours de vulgarisation scientifique dessine, à gros traits, l'une des stratégies d'exposition possibles en indexation.

B- Éléments du discours de vulgarisation dans la stratégie d'exposition des documents en indexation

Rappelons que l'enjeu, en indexation, d'une stratégie d'exposition est le suivant : l'indexation doit pouvoir contraindre l'interprétation que les locuteurs feront des documents. Pour cela, il faut les leur présenter dans un cadre, un contexte d'interprétation. Ce cadre peut être une classification hiérarchique ou - mieux - une classification à facettes. Ce cadre nous semble pouvoir être aussi celui des terminologies spécialisées, du moins tel que le discours de vulgarisation scientifique en fait usage. Le terme de spécialité nous semble pouvoir fonctionner, en indexation, comme il est utilisé en discours par les vulgarisateurs : il doit permettre de rattacher un document à d'autres textes sans que lui-même ne constitue une unité d'information, de même que le terme ne constitue pas une « unité de connaissance » en vulgarisation mais permet de faire référence aux domaines dont il est question. La mise en œuvre du principe d'exposition des documents inspiré du discours vulgarisant pourrait être la suivante.

Les documents peuvent être classés, dans plusieurs classes le cas échéant, en fonction des terminologies qu'ils mettent en œuvre. Le repérage des terminologies, s'il peut se faire de façon automatisée, n'en demande pas moins un travail de validation manuelle : on a pu précédemment montrer en I.2 que les terminologies n'existaient pas en soi mais qu'elles prenaient sens au travers de textes et de locuteurs. Le principe consiste, comme dans les approches précédentes, à regrouper les textes par « domaines » mais le regroupement s'effectue ici sur la base des termes issus des textes eux-mêmes, et non sur la base d'une analyse de contenu à proprement parler, et ne vise pas à un classement unique, mais au contraire multiple. Un texte peut très bien, comme le montre le discours de vulgarisation scientifique, relever de plusieurs domaines différents à la fois. Les termes utilisés pour classer un document dans plusieurs domaines ne sont pas, en outre, nécessairement ceux qui seront proposés comme descripteurs ; nous y reviendrons. Beaucoup d'aspects restent à spécifier pour que ce type de stratégie d'exposition des documents puisse réellement voir le jour. Cependant des tentatives de classification de documents sur la base de leurs terminologies ont déjà été expérimentées¹⁸².

Mais là encore, il nous semble important de ne pas confondre les termes utilisés

vulgarisateur est linguistiquement contrainte.

¹⁸². Voir, sur ce point, le colloque SFBA (Société française de bibliométrie appliquée) de 1995.

pour la répartition des documents en domaines (au sens terminologique) avec les termes utilisés pour circuler entre ces domaines (les descripteurs proprement dits). Les domaines ne sont constitués que pour donner à voir les documents sous un certain angle mais ce n'est pas uniquement sous cet angle-là que peuvent être lus les documents.

Le type de stratégie d'exposition des documents que nous proposons reprend, à rebours, la méthode mise en œuvre par Foucault pour dégager l'espace des pratiques discursives¹⁸³ : on commence par considérer les objets de discours dans le cadre des regroupements académiques dans lesquels ils se donnent pour ensuite élaborer des chemins de traverse entre ces différents regroupements.

C'est ici, comme dans le discours de vulgarisation scientifique, la notion de parcours référentiel effectué par le lecteur qui est mise en avant :

La référenciation adéquate peut être vue comme un processus de construction d'un chemin liant différentes dénominations approximatives qui ne sont pas effacées par le dernier choix. Une conséquence en est que plusieurs tentatives de nomination peuvent être retenues comme adéquates, la correction de l'erreur étant alors utilisée comme une ressource interactionnelle pour invoquer des formulations alternatives.
Dubois et Mondada, 1995, p. 285.

À la problématique du mot juste, dominante en indexation, la vulgarisation scientifique, tout comme les études plus générales sur la désignation, conduisent à envisager une autre problématique, où l'interprétation n'est pas « trouvée », encapsulée qu'elle serait dans une seule et unique unité lexicale, mais où l'interprétation se « trouve » par le biais d'une construction des objets de discours qu'il est nécessaire de guider, de contraindre.

En fin de ce chapitre, nous retrouvons les convergences entre les pratiques documentaire, terminologique et vulgarisante qui nous avaient initialement permis d'établir une première approche du discours documentaire. Mais, désormais, le « continuum », pour reprendre le terme de Jacobi, qui peut être établi entre ces trois pratiques de diffusion des connaissances, peut trouver, au-delà du rapprochement analogique, quelques fondements théoriques. En effet, la notion de monde possible établie par Kripke, qui nous a permis de problématiser la notion de stratégie d'exposition des documents (c'est le monde possible des documents qui doit contraindre la construction des mondes possibles des utilisateurs, et non l'inverse), peut aussi être vue comme un mode de représentation des problématiques documentaires qui légitime le rapprochement opéré avec la pratique vulgarisante, et, par son biais, avec la pratique terminologique.

¹⁸³. Voir précédemment le § II.1.1.

IV- Conclusion du chapitre

Alors que, dans son « premier âge », l'indexation se donnait explicitement sous sa dimension discursive¹⁸⁴, on rencontre aujourd'hui bien plus de difficultés pour la faire apparaître. À maintes reprises dans cette étude, à vouloir capter le discours documentaire, que ce soit sous l'angle de l'exploration des sources ou sous celui de l'exposition des documents, c'est au langage documentaire que nous nous sommes trouvés confrontés, à un langage documentaire qui apparaît en indexation comme l'outil multifonctions par excellence, permettant tout à la fois de sélectionner des sources, d'attribuer des descripteurs et de présenter des documents.

L'un des enjeux de ce chapitre était précisément de montrer que le recours systématique à un même type d'outil lexical ne devait pas masquer les différents types d'opération que l'indexation réalise sur les discours.

Pour ce faire, nous avons multiplié les pistes pour approcher la notion de discours documentaire, particulièrement absente dans la littérature classique sur l'indexation : chacune des approches proposées ici (approche comparative de pratiques ou approche par reformulation de modèles) pourrait, à elle seule, constituer l'objet d'une étude, dont nous n'avons ici proposé qu'une esquisse.

En dépit du caractère suggestif qui caractérise sur bien des points notre démarche, il nous semble que nous avons pu faire apparaître la nécessité - pour la détermination des descripteurs proprement dite, par laquelle se définit habituellement l'indexation - de disposer d'une perspective plus large, qui englobe la problématique de la sélection des sources comme celle de l'exposition des documents. En effet, c'est en appréhendant l'indexation sous cet angle élargi que se dégagent deux systèmes de contraintes qui pèsent sur l'indexation proprement dite, sur le choix *in fine* des descripteurs :

- un premier système de contraintes est représenté par les sources elles-mêmes. Sur ce point, l'approche de Foucault, quelle que soit la difficulté de son utilisation dans un cadre professionnel particulier, montre qu'il existe des règles - des règles discursives - qui président à l'existence des textes. L'indexation ne peut pas ignorer les conditions d'existence des discours, sous peine de les considérer comme un amas d'objets sans cohérence dans lequel on pourrait piocher impunément. La problématique des politiques d'acquisition

¹⁸⁴. Comme le rappelle Escarpit [1991, p. 153], aux premiers temps des bibliothèques, c'est l'*incipit* d'un document qui était retenu pour le « représenter ».

nous semble sur ce point pouvoir bénéficier d'un ancrage théorique tel que le modèle de Foucault en offre une forme générale.

Outre la difficulté de l'entreprise, les conséquences d'un tel choix théorique sur l'approche du descripteur lui-même ne sont pas négligeables : en effet, ce que suggère la démarche de Foucault, comme le met bien en valeur Milner¹⁸⁵, c'est que toute pratique discursive est saisie dans un réseau de termes spécifiques qui constitue ses propres objets de discours. Par conséquent, une pratique d'indexation qui s'inscrirait dans le modèle théorique des formations discursives, s'appliquerait nécessairement à choisir des descripteurs qui soient issus des discours eux-mêmes. L'affectation d'autres mots, issus d'autres « histoires » dans les termes de Foucault, ruinerait complètement l'intérêt de la démarche. Autrement dit, l'indexation ne peut être dans un tel cadre qu'une **extraction de mots de discours** ;

- le second système de contraintes qui pèse sur l'indexation est, lui, lié à la problématique de l'identité des objets en indexation. Dans la première partie de cette étude, nous avons examiné le modèle implicite sur lequel reposait la croyance en la stabilité référentielle des objets en indexation. Dans cette seconde partie, nous avons cherché à redéfinir cette question en la posant dans le cadre théorique des mondes possibles établi par Kripke. Problématisée dans ce cadre, la question de la stabilité référentielle ou encore de l'identité des objets s'exprime clairement en termes de contraintes : qui, des indexeurs ou des utilisateurs, définit les objets à partir desquels vont pouvoir être examinées les différentes propriétés ? Du fait qu'il s'agit de donner à manipuler des objets dont il est question dans les textes d'une collection documentaire, seuls les indexeurs sont à même de déterminer l'ensemble des objets identifiables. De là, la nécessité pour l'indexation de contraindre le parcours interprétatif de l'utilisateur. De là aussi la nécessité de repenser la morphologie du descripteur : pour pouvoir établir une relation entre univers de référence, les descripteurs doivent présenter un certain degré de « rigidité ». En cela, les descripteurs doivent être nécessairement des termes.

¹⁸⁵. Milner 1989, p. 66.

Dans le cadre d'une approche discursive de l'indexation, la problématique du descripteur se laisse approcher sous une forme sensiblement différente que celle proposée dans les normes ou la littérature classique ; nous spécifierons, dans le prochain chapitre, les deux caractéristiques dégagées ici du descripteur : unité extraite du discours et terme.

À l'issue de ce chapitre, il apparaît que le processus de l'indexation peut être appréhendé sans qu'il soit nécessaire de recourir à la notion de langage documentaire, à la notion d'uniformisation lexicale. La notion de discours documentaire, définie comme espace d'organisation spécifique des documents - spécifique par son mode d'exploration des sources et par son mode d'exposition des documents -, si elle n'a pas été ici entièrement définie, apparaît non pas comme ce qui homogénéise les documents, mais plutôt comme ce qui permet de maintenir, ensemble et accessible, une diversité de documents hétérogènes.