

Chapitre II - La question du lexique en indexation

La question du lexique¹ en indexation constitue, de façon typique, pourrait-on dire, une zone de tension entre modèle d'utilisation et modèle de fonctionnement de la langue.

Les deux modèles se disputent en effet les « mêmes objets » : les « mots »² - descripteurs* dans le premier modèle, unités lexicales* dans le second - appréhendés selon des points de vue, des modes de saisie radicalement différents :

- le discours classique refuse la notion de lexique, au nom de son inorganisation et de son ambiguïté³. Il y substitue la notion de langage documentaire, dont la visée pragmatique est censée corriger les imperfections relevées. Le mode de saisie de ses unités se veut déconnecté des notions intuitive de « mot » et linguistique d'« unité lexicale »⁴ ;
- l'approche linguistique refuse, elle aussi, la notion intuitive de lexique, notamment parce qu'y est privilégié un point de vue « externe » sur le lexique (vu comme une instance d'enregistrement d'expressions linguistiques « données ») au détriment d'un point de vue « interne »⁵. Considéré d'un point de vue interne, le lexique constitue un artefact⁶, susceptible de faire l'objet d'une théorie, dans laquelle les unités lexicales peuvent trouver des formes de régularité et d'organisation⁷.

¹. Dans cette première occurrence, le lexique est entendu dans son acception « intuitive » comme l'ensemble des mots d'une langue.

². On emploie ici encore, pour le début de ce chapitre, une notion intuitive, celle de « mot », sans chercher à la définir. On emploiera également le terme de « forme lexicale » pour renvoyer à cette même notion intuitive de « mot ».

³. Par exemple, la norme AFNOR Z 47-100 (1981), p. 185 : « Un thésaurus ne doit être confondu ni avec un lexique, ni avec un index, ni avec un dictionnaire. [...] Le vocabulaire constituant le thésaurus doit être **non ambigu** pour que les mêmes termes de ce vocabulaire identifient systématiquement les mêmes concepts ; **structuré** pour assurer une meilleure définition de chaque terme et pour permettre des recherches à différents degrés, de généralité ou de spécificité ». (C'est nous qui soulignons).

⁴. Chaumier 1978, p. 29-30 : la « signification » des descripteurs est fixée par la structure des thésaurus notamment grâce aux relations hiérarchiques qui « ne sont pas nécessairement basées sur des critères sémantiques mais aussi sur des considérations pragmatiques résultant des besoins de la recherche documentaire ».

⁵. Marandin 1992a [présentation], p. 6.

⁶. Rastier 1994, p. 28.

⁷. Marandin 1992a [présentation], p. 6 : « La question d'une théorie autonome du lexique se pose dès lors que l'on suppose que les caractères d'inorganisation, d'irrégularité du lexique, d'idiosyncrasie

La notion intuitive de lexique est donc critiquée dans les deux approches mais selon des points de vue différents : en raison de contraintes d'utilisation dans le modèle de l'indexation, en raison de critères scientifiques dans le modèle linguistique. Cependant, dans les deux cas, on substitue pareillement, à la notion intuitive de lexique, la notion de modèle : modèle de fonctionnement en linguistique, modèle d'utilisation en indexation, mais modèle d'utilisation qui, dans le discours classique, se donne pour ou qui dit se passer d'un modèle de fonctionnement.

Or, en étudiant les différentes fonctions attribuées au descripteur, on constate que, loin de prendre ses distances avec un modèle de fonctionnement de la langue, le modèle d'utilisation du lexique en indexation repose sur les conceptions les plus traditionnelles et les plus courantes du lexique. On peut en effet relever les formes d'une *doxa* linguistique⁸ qui constituent implicitement l'arrière-plan théorique du modèle d'utilisation du lexique en indexation. Cet arrière-plan théorique révèle, sur un plan pratique, des limites et des insuffisances, qui sont alors imputées, par les professionnels, à la langue elle-même ; or c'est à une certaine vision du lexique qu'elles devraient plutôt être rapportées. L'arrière-plan théorique implicitement présent dans le discours classique présente en outre l'inconvénient de ne pouvoir rendre compte des propriétés des unités lexicales réellement en jeu dans l'indexation. Comme nous le montrerons, la *doxa* linguistique joue, dans la description classique de l'indexation, un rôle fortement opacifiant.

Pour tenter de démêler ces différents types et niveaux de représentations linguistiques en jeu dans cet aspect de l'utilisation des « mots » en indexation, nous procéderons de la façon suivante :

- à partir du discours de la norme⁹, qui constitue le cœur des approches classiques, nous étudierons le modèle du lexique en indexation sous ses deux aspects : le modèle d'utilisation du lexique (les différentes fonctions du descripteur) et le modèle de fonctionnement sous-jacent (la *doxa* linguistique) ;
- nous examinerons ensuite les fonctions attribuées au descripteur à la lumière d'un autre modèle de fonctionnement, lui explicitement linguistique, des unités lexicales : nous pourrions distinguer, parmi ces différentes fonctions, celles qui relèvent d'effets propres à une utilisation et celles qui relèvent de principes propres à la langue ;

des entrées sont davantage liés au point de vue adopté qu'à l'objet que l'on discerne sous la construction théorique, à savoir le matériel lexical propre à une langue ».

⁸. Rappelons que, dans cette recherche, la notion de « *doxa* linguistique » renvoie à l'ensemble des représentations que tout sujet parlant se forge sur la langue et le langage.

⁹. Norme AFNOR Z 47-102 (Principes généraux pour l'indexation des documents) 1978, reproduit en annexe 1.

- suite à ce déplacement de modèles de fonctionnement du lexique, nous reformulerons le modèle d'utilisation du lexique en indexation. La distinction des faits et des effets nous conduira à mettre à distance le rôle des « mots » en indexation et ce à deux niveaux : au niveau de la description du processus de l'indexation, dont on proposera une nouvelle formulation, et au niveau de la définition des descripteurs, dont on réévaluera les fonctions.

I- Le modèle du lexique en indexation

Si, comme nous l'avons précédemment relevé, il n'est pas question, à proprement parler, de lexique dans le discours classique, il y est en tout cas question des « mots », et plus précisément du rôle que ceux-ci sont censés jouer en tant que descripteurs¹⁰. Il nous paraît important de ne pas masquer, ne serait-ce qu'à travers la notion confuse de « mot » dans un premier temps, la dimension lexicale des descripteurs puisque, malgré l'idéal qu'elle se donne¹¹, l'indexation continue d'emprunter à la langue son matériel lexical, alors même que certains chercheurs ont pu proposer des langages documentaires « véritablement » artificiels¹².

C'est en fonction de cette approche du descripteur que l'on dégagera les deux niveaux du modèle du lexique en indexation : niveau de l'utilisation et niveau des représentations linguistiques sous-jacentes.

I.1- Modèle d'utilisation du lexique en indexation

En reprenant l'esprit de la norme¹³, on peut dire que l'indexation d'un document* se traduit par l'attribution à un document d'un ou de plusieurs mots, ces mots entretenant avec le document un rapport de deux types :

10. Notre approche nous conduit à considérer la notion de descripteur dans une acception plus large que celle que lui accorde la norme (voir glossaire). Le descripteur sera à entendre dans cette recherche, sauf indication contraire, comme toute forme lexicale utilisée pour indexer, que cette forme lexicale fasse partie ou pas d'un langage documentaire. En effet, la distinction entre indexation libre (indexation par choix libre de termes) et indexation contrôlée (indexation par choix dans une liste close) que sous-tend la définition normative du descripteur ne peut être prise en compte à ce stade de l'étude : elle implique en effet une distinction que nous discutons, qui consiste à poser une différence de « nature » entre mots de la langue et descripteurs là où nous proposons de voir une distinction d'« utilisation ».

11. C'est chez un linguiste que l'on trouve, selon nous, la meilleure expression de cet idéal : « Ce qui est souvent admis comme résultat idéal pour un automate dans ce domaine [documentaire] c'est un texte en langue naturelle vidé des supports linguistiques des chaînes et présentant en leur lieu et place des identificateurs symboliques, des indices. Soit à peu près : remplacer, en utilisant des nombres ou des lettres, les expressions d'un texte par le symbole de ce qui est désigné », Corblin 1987, p. 15.

12. C'est toute la réflexion de Gardin et de son équipe menée dans les années 1950 sur le langage « mécanographique » dans le cadre du système SYNTHOL (*SYNTagmatic Organization Language*). Leur recherche portait sur la façon d'« obtenir une analyse sémantique des objets qui soit indépendante de leurs noms dans les langues naturelles » ; ils proposaient, pour ce faire, la constitution de véritables codes, fondés exclusivement sur des symboles (lettres ou chiffres), à l'exclusion de toute forme lexicale. On peut se reporter à Bely et *al.* 1970, Gardin 1967 et Gardin 1974 ; dans Gardin 1991, on trouve une synthèse et une évaluation d'une partie des travaux menés dans le cadre du système SYNTHOL.

13. Voir les deux fonctions de l'indexation : « permettre la recherche efficace des informations contenues dans un fonds documentaire » et « indiquer brièvement, sous forme concise, la teneur du document ». Norme AFNOR Z 47-102 1978, p. 225.

- un rapport de représentation : l'indexation fournit la représentation du contenu d'un document. Cet aspect correspond, dans le texte de la norme, à la notion d'indication de la « teneur » d'un document ;
- un rapport de catégorisation : l'indexation indique l'appartenance d'un document à un (ou plusieurs) ensemble(s) de documents jugés semblables. Cet aspect correspond, dans le texte de la norme, à la notion d'accès non à un document mais « aux informations contenues dans un fonds documentaire ».

À partir de cette première paraphrase, on dira que le descripteur se voit attribué, dans la norme, deux fonctions : il est à la fois une unité de représentation du contenu d'un document et une unité d'accès à l'information d'un ensemble documentaire.

L'indexation pose donc le descripteur dans un rapport à deux niveaux distincts, sans que l'on sache comment s'établit le passage de l'un à l'autre : pensé dans ses rapports avec un document, le descripteur s'approche en termes de contenu ; pensé dans ses rapports avec un fonds documentaire, il est décrit en termes d'accès.

Pour essayer de préciser ces deux fonctions du descripteur et le lien éventuel qu'elles entretiennent, nous commencerons par les examiner chacune séparément.

I.1.1- Le descripteur en tant qu'unité de représentation du contenu d'un document

On peut comprendre ce que le discours normatif entend par les notions de « contenu » et de « représentation » en examinant la description du processus de l'indexation qu'il donne.

L'indexation y est décrite comme un processus réalisé en deux phases :

- une phase d'analyse de contenu, dite phase d'analyse conceptuelle où s'effectue « la reconnaissance des concepts contenant l'information dans les documents à indexer » ;
- une phase de représentation dite phase de traduction correspondant à la « représentation de ces concepts dans le langage documentaire ».

Cette appréhension des notions de « contenu » et de « représentation » détermine fortement les caractéristiques dont doit être pourvu le descripteur : nous les précisons ci-après.

A- La notion de « contenu » : le descripteur comme expression linguistique de concept

Il apparaît que la notion de contenu est, de façon générale en indexation, de nature conceptuelle : qu'il s'agisse de caractériser le contenu d'un document ou le contenu d'un descripteur, c'est toujours à la notion de « concept » que l'on se réfère. S'il est cependant plutôt question de « notion » pour qualifier le contenu d'un descripteur¹⁴ et de « concept » pour qualifier celui d'un texte, les deux termes sont en fait unifiés par le biais d'une même définition¹⁵.

Le descripteur, comme élément d'un langage documentaire ou comme élément de description d'un document est, dans tous les cas, l'expression linguistique d'un concept. Cette caractéristique repose sur les présupposés suivants :

- les concepts sont donnés comme pré-existants, c'est-à-dire comme existant préalablement à toute formulation linguistique (le descripteur est un terme qui « renvoie » à un concept) comme à toute analyse documentaire (les concepts sont « reconnus » et « extraits » d'un document) ;
- les concepts sont donnés comme stables : les concepts d'un document sont les mêmes que ceux des descripteurs qui sont les mêmes que ceux des requêtes documentaires¹⁶ ; autrement dit, la formulation linguistique d'un concept n'obère pas son appréhension¹⁷ ;
- les concepts de l'indexation sont des concepts « simples » qui doivent correspondre, idéalement, à des mots « simples ». L'assimilation concept/mot se joue en indexation jusque dans la détermination des formes du descripteur : la forme linguistique du descripteur doit être « décomposée » en fonction d'une « décomposition » conceptuelle¹⁸.

¹⁴. Voir la définition normative du langage documentaire : « langage artificiel constitué de **représentations de notions** et de relations entre ces notions et destiné, dans un système documentaire, à formaliser les données contenues dans les documents et dans les demandes des utilisateurs ». Vocabulaire de la documentation, 1987 (c'est nous qui soulignons).

¹⁵. La norme AFNOR [Z 47-102 (1978), p. 231] assimile « notion » et « concept » par une définition commune : « toute unité de pensée ».

¹⁶. La croyance en une stabilité du concept se situe au cœur du projet des langages documentaires, cf. norme Z 47-100 1981, p. 3 : « Le thésaurus permet donc de traduire en termes d'indexation ou en termes de recherche **tout concept devant entrer ou sortir d'un système documentaire donné** ». C'est nous qui soulignons.

¹⁷. Cette idée apparaît également de façon claire chez Ranganathan [1967], qui distingue, lui, trois phases dans l'indexation (« Idea Plane », « Verbal Plane », « Notational Plane ») : « **subject analysis takes place in an Idea Plane which words we use to express our ideas is not important** ». C'est nous qui soulignons.

¹⁸. Cf. norme 47-100 1981, p. 186 : dans le cas d'« expressions complexes », il faut décomposer de façon parallèle mots et concepts. « Il existe deux manières d'analyser les termes complexes pour y reconnaître les notions simples :

Pour être une unité de représentation du contenu d'un document, le descripteur doit donc fonctionner comme l'expression linguistique d'un concept pré-existant, stable et simple. La norme ne précise pas si c'est cette expression linguistique du concept ou le concept lui-même qui constitue une information, une partie de l'information d'un document.

B- La notion de « représentation » : le descripteur comme expression linguistique

La notion de représentation relève, dans le discours classique, de deux plans différents :

- **sur un plan lexical**, il s'agit de représenter un concept, de « mettre en mots » les concepts stables et simples pré-existants. Cette opération s'effectue sur le mode de la « traduction », elle-même de deux types : une « traduction » de type conceptuel (concept -> mot), une « traduction » de type interlingual (mot₁ -> mot₂¹⁹) ;
- **sur un plan textuel**, il s'agit de représenter un texte, c'est-à-dire de réduire le texte du document à indexer. La réduction porte elle aussi sur deux plans²⁰ : celui de la « forme » (l'indexation, c'est le texte original en plus petit) et celui du « fonds » (l'indexation ne dit pas tout du texte, seulement ce qui est le plus essentiel²¹).

-
- l'analyse sémantique, exemple : oxycoupage = découpage+oxygène ;
 - l'analyse morphologique (syntaxique), exemple : psychologie des animaux = psychologie+animal ».

19. Où « mot₁ » est le premier terme venant à l'esprit de l'indexeur pour exprimer le concept et « mot₂ » le mot du langage documentaire, le descripteur, s'approchant au mieux du « mot₁ ».

20. Ces deux aspects de la réduction ne sont pas toujours distingués. On ne sait pas toujours, chez certains auteurs, si l'indexation opère une condensation de tout le texte (avec les pertes « informationnelles » liées au principe de la condensation) ou une focalisation sur certaines parties du texte. On trouve, en effet, des formulations diverses : l'indexation relève chez les uns d'une représentation exhaustive du texte d'origine (Dewèze 1993, p. 165 : « moyen d'expression réduite du texte intégral » ; Bertrand-Gastaldi, 1986, p. 4 : « opération qui vise à substituer au texte de départ un texte d'arrivée, plus court et mieux organisé, plus économique à mémoriser et à manipuler »). Chez d'autres auteurs, l'indexation relève d'une représentation sélective du texte (Chaumier, 1988 : « expression plus ou moins condensée des caractéristiques d'un document » ; de même la norme Z 47-102, qui préconise une « sélection des concepts », voir note 21 suivante).

21. Outre que l'indexation n'explique pas toujours ce qu'elle choisit de dire d'un texte, les critères de sélection sur lesquels elle repose paraissent difficiles à formaliser : tout dépend de la « fonction », c'est-à-dire de la fameuse finalité donnée à l'indexation. Sur ce point, voir, par exemple, la norme Z 47-102, 1978, p. 227 : « Tous les concepts identifiés par l'indexeur comme représentatifs de l'information contenue dans le document ne sont pas nécessairement retenus pour l'indexation. En effet, l'indexation doit essentiellement être adaptée à sa fonction propre dans le système où elle est utilisée ». C'est nous qui soulignons.

Ces deux aspects de la représentation en indexation reviennent à attribuer des caractéristiques particulières au descripteur :

- le descripteur doit fonctionner comme une unité stable (mais on ne sait pas s'il s'agit de la même stabilité que celle reconnue par ailleurs au concept) : il doit en effet maintenir le « concept » d'origine à travers les différentes formes linguistiques qui peuvent servir à l'exprimer ;
- le descripteur doit fonctionner comme un condensateur textuel : être une expression du texte à la fois réduite du point de vue de la forme et synthétique du point de vue de son « sens ».

Pour pouvoir fonctionner comme une unité de représentation du contenu d'un document, un « mot » donc doit relever simultanément :

- du concept : il doit être une expression conceptuelle, pouvoir exprimer un concept ;
- du symbole : il doit être une expression stable, pouvoir exprimer toujours le même concept ;
- du texte : il doit être une expression textuelle, pouvoir exprimer un texte ou certaines de ses parties.

En plus de ces trois caractéristiques liées à la première de ses fonctions s'en ajoutent d'autres, liées, elles, à la fonction du descripteur comme accès à un ensemble documentaire.

I.1.2- Le descripteur en tant qu'accès à un ensemble documentaire

Le discours classique détaille fort peu cette fonction de l'indexation (permettre la recherche d'information) et cet aspect du descripteur (accès à l'information d'un ensemble documentaire).

On retrouve là la dissolution de l'objet « indexation » propre aux approches classiques : insensiblement le discours sur l'indexation se tait et cède la place à celui sur la recherche documentaire.

Cependant, il y a implicitement, dans le discours classique, l'idée que l'accès à l'information d'un fonds documentaire se fait sur la base des unités de représentation de contenu de chacun des documents qui le constituent²².

²². On trouve chez Menon une formulation explicite de ce lien entre représentation du contenu et

Il revient donc bien à un modèle de l'indexation (et non à un modèle de la recherche documentaire) d'expliquer comment peut se réaliser ce double saut :

- saut du document à une pluralité de documents ;
- saut du « contenu » à l'« information ».

Il importe donc de déterminer les caractéristiques du descripteur en fonction de ces deux rôles.

Le modèle d'utilisation du lexique en indexation ne dit rien de la fonction du descripteur comme accès à l'« information » contenue dans plusieurs documents différents. Dans ce paragraphe, nous laissons donc ce point de côté ; nous le reprendrons dans le cadre d'un autre modèle. En revanche, il nous fournit quelques éléments sur le fonctionnement du descripteur comme accès à un document d'une part et comme accès à plusieurs documents d'autre part.

A- L'accès à un document : le descripteur comme une expression linguistique autonome

Le discours classique ne relève ni ne commente le fait que les descripteurs retenus pour indexer un document puissent être lus, interprétés et utilisés, seuls, détachés du document et/ou du langage documentaire d'où ils proviennent. Il y a là une évidence que l'ensemble des mots issus de l'indexation doit exprimer quelque chose pour un utilisateur (mais quoi ? le contenu d'un document ? de l'information ? un sens ?).

De cette évidence, qui ne va pas de soi²³, on peut néanmoins dégager une caractéristique dont doit être doté le descripteur : il doit révéler une certaine autonomie, lui permettant de fonctionner seul. Nous aurons à déterminer le type d'autonomie dont le descripteur doit être pourvu : conceptuelle ? référentielle ? sémantique ?

B- L'accès à plusieurs documents : le descripteur comme relais textuel

La norme suggère que le descripteur joue un rôle non seulement par rapport à un document mais aussi par rapport à plusieurs documents : il doit permettre, au sein d'un fonds documentaire, d'établir des liens, d'effectuer des rapprochements, de constituer des ensembles. Un même descripteur doit donc pouvoir être affecté à des documents différents, soit parce que les « concepts qui s'y trouvent » sont jugés

accès à l'information : « Nous nous intéresserons surtout ici aux unités susceptibles de permettre l'accès à l'information par des éléments tirés de son contenu thématique », Menon 1988, p. 146.

²³. On peut toujours trouver un « sens » à une série de mots : c'est ce que David et Plante [1990a] ont proposé d'appeler le « syndrome de la signification garantie ». En revanche, on ne peut jamais être sûr que ce sens corresponde à celui du texte dont ces mots sont extraits ou pour lequel des mots ont été choisis.

semblables, soit parce que les « concepts qui s'y trouvent » sont posés comme équivalents au concept que « représente » le descripteur.

Il y a donc dans la notion de descripteur celle de classe d'équivalence : est-ce une classe d'équivalence lexicale (de « mots ») ? Est-ce une classe d'équivalence documentaire (de « choses ») ?

La norme n'est pas très claire sur ce point. Si elle dit assez nettement qu'il ne s'agit pas de classe d'équivalence lexicale, elle ne dit pas explicitement qu'il s'agit de classe d'équivalence documentaire. En effet, le descripteur y est défini comme étant un « terme préférentiel », c'est-à-dire un terme retenu de préférence à d'autres au sein d'un ensemble de termes jugés équivalents. Dans la norme, le jugement d'équivalence ne s'établit pas en fonction du « sens » des termes²⁴ mais en fonction de leur emploi supposé²⁵, ce qui renvoie à nouveau le problème dans le camp de l'utilisateur et de la recherche documentaire²⁶.

Même si elles sont peu détaillées dans le discours classique, la présence de la notion de classes d'équivalence - plutôt de type documentaire que lexical -, et celle, corollaire, de synonymie, signalent qu'un descripteur doit être à même d'indiquer un jugement de ressemblance entre documents : on dira qu'en cela le descripteur doit jouer un rôle de relais entre textes, de relais textuel.

I.1.3- Caractéristiques du descripteur : récapitulatif

Sous forme de récapitulatif, on rappelle comment les deux fonctions attribuées à l'indexation dans le texte de la norme déterminent celles du descripteur et comment ces fonctions du descripteur peuvent se laisser traduire en caractéristiques dont il doit être doté²⁷ :

²⁴. Chaumier 1978, p. 33 : « La notion de synonymie est utilisée de façon extensive dans les thésaurus sous la forme de la synonymie documentaire afin de regrouper sous un seul descripteur plusieurs termes considérés comme voisins bien que de signification sémantique différente ».

²⁵. Norme Z 47-100 (1981), p. 189 : « L'établissement des relations d'équivalence doit être gouverné par le principe simple suivant : lorsque tout document indexé à l'aide du terme A doit être pris en considération pour toute demande indexée à l'aide du mot B (et réciproquement), les termes A et B sont synonymes documentaires (soit vrais synonymes soit quasi-synonymes). Il suffit alors de choisir l'un des termes A ou B comme descripteur, l'autre étant non-descripteur ».

²⁶. *Ibid.*, p. 190 : « Le choix des termes à utiliser comme descripteurs à l'intérieur des classes d'équivalence doit obéir à des critères à établir à partir des besoins de la majorité des utilisateurs potentiels ».

²⁷. La numérotation adoptée ci-dessous ne traduit aucune idée d'ordre (dans le processus de l'indexation par exemple) ; elle est utilisée pour faciliter les renvois ultérieurs.

- **fonction 1 de l'indexation** : « indiquer brièvement, sous forme concise, la teneur du document »
fonction 1 du descripteur : le descripteur comme unité de représentation du contenu d'un document :
 - > caractéristique.1 : le descripteur comme expression linguistique d'un concept pré-existant, stable, unitaire (« simple ») ;
 - > caractéristique.2 : le descripteur comme expression linguistique stable ;
 - > caractéristique.3 : le descripteur comme condensateur textuel.
- **fonction 2 de l'indexation** : « permettre la recherche efficace des informations contenues dans un fonds documentaire »
fonction 2 du descripteur : le descripteur comme accès à un fonds documentaire :
 - > caractéristique.4 : le descripteur comme expression linguistique autonome ;
 - > caractéristique.5 : le descripteur comme relais textuel.

Dans ce paragraphe, on a tenté de montrer comment s'organisait le modèle d'utilisation des mots en indexation. À partir des fonctions de l'indexation se déduisent celles du descripteur. Ces fonctions font l'objet d'une description, qui, pour ne pas être toujours complète, permet néanmoins à l'indexation de se constituer en pratique²⁸. Cette description revient à attribuer, de façon explicite et implicite, des caractéristiques au descripteur : c'est lors de ce passage des fonctions aux caractéristiques du descripteur que le discours classique emprunte à un modèle courant du lexique ses présupposés, comme nous le verrons ci-après.

Il semble en effet que, pour être utilisés de la façon dont l'indique la norme, les descripteurs n'ont pas nécessairement besoin d'être pourvus des caractéristiques qui leur sont attribuées explicitement et implicitement : il y a, dans l'approche du descripteur dans le discours normatif, l'adoption implicite de représentations linguistiques qui, pour être communes à l'ensemble des sujets parlants, ne s'en révèlent pas moins à la fois inadaptées et limitées pour décrire les faits d'indexation.

²⁸. Dans l'acception courante du terme, « manière concrète d'exercer une activité ».

I.2- Modèle de fonctionnement implicite du lexique en indexation

Le modèle d'utilisation du lexique en indexation, précédemment exposé, repose sur l'articulation de deux types de représentation linguistique des plus courants, représentations que les sciences du langage ont pu remettre en question :

- le premier type concerne la nature même de la langue, appréhendée sous le seul angle de ses formes lexicales : c'est ce que l'on nommera le modèle lexicaliste ;
- le second type, corollaire du premier, concerne la fonction du langage comme instrument de communication : ce présupposé relève de ce qu'il est convenu d'appeler un modèle objectiviste du langage²⁹.

I.2.1- Un modèle « lexicaliste » de la langue

A- Approche du modèle lexicaliste

On aura pu remarquer, dans le précédent paragraphe, que l'ensemble du processus de l'indexation passe par les mots et ne met en jeu que des mots : les concepts d'un texte sont des mots, les concepts des descripteurs sont des mots, les ressemblances entre textes se font sur la base de mots, etc.

Les mots en indexation assimilent donc, par le biais des « descripteurs », des objets de nature différente (notamment les objets cognitifs et les objets textuels³⁰).

On qualifiera cette approche de « lexicaliste » pour signifier qu'elle ne pense la langue que sous un seul rapport : celui des formes lexicales.

B- Critiques du modèle lexicaliste

L'un des enjeux de la linguistique contemporaine a été précisément de montrer que la langue était constituée de plusieurs dimensions³¹ et que les formes lexicales ne pouvaient constituer l'accès unique et exhaustif au « sens » ou encore au « contenu » des textes. L'une des premières tâches de la linguistique a donc consisté à démonter la vision de la langue comme nomenclature, comme répertoire. C'est Saussure qui le premier a mené la critique³², relayée depuis par des chercheurs de différents courants, structuralistes et non structuralistes. Plusieurs

29. Dubois 1995.

30. Comme le fait remarquer Dubois 1995, p. 88-89.

31. Par exemple, Milner 1989.

32. Sur la base de départ suivante : « Pour certaines personnes, la langue, ramenée à son principe essentiel, est une nomenclature, c'est-à-dire une liste de termes correspondant à autant de choses [...]. Cette conception suppose des idées toutes faites préexistant aux mots », Saussure 1973, p. 97.

aspects du modèle de la langue comme nomenclature ont été diversement discutés dans le champ linguistique, notamment celui du sens des mots.

En effet, le modèle de la langue-répertoire conduit à envisager la dénomination* sous la forme d'un baptême, selon une procédure canonique dont on peut voir, comme le propose Mounin³³, la première expression dans la Bible : « Et Dieu nomma la lumière Jour, et les ténèbres Nuit [...] Et Dieu nomma l'étendue Cieux », etc. Dans ce modèle, les mots n'ont pas de signification qui leur est propre : ils sont perçus comme des symboles, des conventions interchangeables. Or, quelle que soit leur approche, les linguistiques contemporaines accordent toutes une place à la signification lexicale, comprise comme l'une des dimensions susceptibles de donner une « individualité » singulière aux « mots », permettant de les distinguer les uns des autres³⁴.

C- Marques du modèle lexicaliste en indexation

Pour discutée qu'elle puisse être dans le champ linguistique, cette approche de la langue comme nomenclature et de la dénomination comme baptême n'en constitue pas moins le modèle de représentation courant de la langue, un type de représentation linguistique standard, pourrait-on dire³⁵. Que le modèle de l'indexation l'adopte n'est donc pas en soi étonnant et n'aurait rien de gênant s'il ne conduisait à mener une description à la fois partielle et insuffisante de la fonction et du rôle des mots en indexation. En effet, nous pourrions montrer (en II) que l'ensemble des fonctions attribuées au descripteur ne relèvent pas toutes de leur nature lexicale, qu'il faut invoquer d'autres mécanismes langagiers, d'autres niveaux, pour décrire les faits d'indexation.

L'une des marques les plus perceptibles du modèle lexicaliste en indexation se trouve dans le mode de dénomination qu'elle met en œuvre. L'indexation, qui ne pose pas la question de la signification lexicale, emprunte le modèle ancestral du baptême. Ainsi, lors de la constitution d'un langage documentaire, procède-t-on à un nouveau baptême des mots de la langue « naturelle » : on décide, par exemple³⁶, de « nommer », par « circulation verticale », « escalier mécanique », « ascenseur », et « monte-charge ». Ce mode de nomination, s'il est caractéristique des unités linguistiques de type « nom propre »³⁷, est appliqué

³³. Mounin 1963, p. 25-26.

³⁴. Même si, comme le relève Marandin, il n'y a aucun consensus sur ce qu'est la signification ; Marandin 1992a [présentation], p. 8 : « Alors qu'il n'y a aucun consensus sur le mode d'individuation par la signification (parce qu'il n'y a pas de consensus sur ce qu'est la signification), il semble que tout le monde s'accorde pour voir dans la signification un principe absolu d'individuation ».

³⁵. Ou plutôt devenu standard au fil des siècles, voir Mounin 1963, p. 25-27.

³⁶. L'exemple est repris de Chaumier 1978, p. 34.

³⁷. Voir ci-après le chapitre III, § III.2.1.

en indexation à tous les types de nom, y compris les noms communs ; nous y reviendrons.

I.2.2- Un modèle objectiviste du langage

A- Approche du modèle et rappel des critiques dont il fait l'objet

Dans le modèle objectiviste, le langage est appréhendé comme un instrument au service d'une fonction : la communication. Il est en cela un mode de codage, un code.

Il ne manipule qu'un seul type de donnée : l'information, qui existe en dehors de la langue et du langage, prête à être codée. L'information étant ce qui peut être transmis, elle peut être de nature différente : concept, sens, signification, connaissance, etc., ne sont pas, dans ce modèle, distinguables.

La notion d'information ne permet pas donc pas d'établir de différences entre les mots et les textes : les textes sont appréhendés comme des ensembles de mots, des supports d'information.

Les trois principaux présupposés du modèle objectiviste ont été largement remis en question par les théories linguistiques³⁸ qui posent au contraire que³⁹ :

- (i) la langue n'est pas un instrument ;
- (ii) si elle permet de communiquer, cette utilisation n'est qu'une parmi d'autres⁴⁰ ;
- (iii) la langue n'est pas un code : son « contenu » ne peut à ce titre en être dissocié.

Les critiques dont le modèle objectiviste a pu faire l'objet n'empêchent pas, bien entendu, que ses présupposés restent valides pour les sujets parlants et ce d'autant plus que ce modèle de la langue comme moyen de communication bénéficie d'une longue et estimable tradition⁴¹. Là encore, il est tout à fait naturel que l'indexation emprunte les présupposés de ce modèle ambiant ; mais là encore, l'inscription dans ce modèle mérite d'être révisée dès lors que son incidence dépasse la simple référence culturelle implicite. En effet, les présupposés de ce modèle constituent, pour la pratique de l'indexation, d'importants facteurs d'opacification qui tendent à

38. Par exemple Chomsky [1975] 1981, mais il n'est pas, loin de là, le seul.

39. On reprend la synthèse des critiques que propose Rastier (1994, n. 1, p. 17-18).

40. « M. Halle a raison de s'élever contre l'attitude de ceux qui, donnant à la formule " une langue est un instrument de communication ", l'interprétation extrapolée " une langue est un instrument parfait de communication " et, constatant qu'il n'en est rien, en prenant l'exact contre-pied, en une formule plus contestable encore : la langue n'est pas un moyen de communication ». Kerbrat-Orrechioni 1980, p. 12.

41. Que la langue soit un instrument de communication et qu'à ce titre elle doive être un parfait moyen de communication se situe par exemple au cœur du projet de « langue universelle » de Descartes ; il y a, dans l'histoire, bien d'autres exemples de cette quête mythique d'une langue parfaite ; on peut sur ce point consulter Eco 1994.

l'enfermer dans une impasse.

B- Les marques du modèle objectiviste en indexation

Le modèle objectiviste, adopté « inconsciemment » en indexation, se trouve exprimé dans le discours classique sous deux formes différentes :

- (i) sous une forme positive : les présupposés du modèle objectiviste permettent de définir les caractéristiques du descripteur ;
- (ii) sous une forme négative : les limites que l'indexation rencontre dans l'utilisation qu'elle veut faire de la langue tiennent des limites propres aux présupposés du modèle objectiviste.

(i) Les caractéristiques du descripteur décrites dans les termes du modèle objectiviste

Certaines des caractéristiques du descripteur que nous avons précédemment dégagées relèvent typiquement du modèle objectiviste, notamment les caractéristiques 1, 2 et 3 :

- le fait que le descripteur soit posé comme l'expression linguistique de concepts pré-existants relève du présupposé objectiviste concernant l'extériorité du « contenu » véhiculé par la langue ;
- l'appréhension du descripteur comme une expression linguistique stable tient, elle, de l'hypothèse de la langue comme code ;
- la caractéristique du descripteur comme condensateur textuel s'inscrit, elle, dans le cadre d'une approche élargie et extensible de la notion d'information qui, dans le modèle objectiviste, annule les différences de niveaux texte/mot.

Quelle conclusion tirer du constat que certaines des caractéristiques du descripteur reposent sur des représentations linguistiques non valides d'un point de vue théorique sur la langue ?

Nous sommes devant l'alternative suivante :

- soit ces caractéristiques décrivent de façon adéquate le descripteur : elles devront être maintenues, sous une autre forme, dans un autre cadre, si on cherche à les constituer en fondements ;
- soit ces caractéristiques sont essentiellement liées aux formes du modèle objectiviste : elles disparaissent alors avec lui.

Nous montrerons que, sur les cinq caractéristiques du descripteur dégagées

précédemment, trois ne se justifient que dans le cadre objectiviste. On verra en effet que les caractéristiques du descripteur comme expression de concept, comme expression stable et comme condensateur textuel, caractéristiques propres à réaliser la fonction 1 de représentation du contenu d'un document, ne rencontrent aucune des propriétés de langue telles qu'un modèle linguistique peut les décrire. En revanche, ces trois caractéristiques peuvent, dans le cadre d'une théorie de la langue, recevoir un autre type d'explication, en termes d'effet d'interprétation. En ce sens, on dira que la première des fonctions de l'indexation - la représentation du contenu d'un document - relève d'un effet d'interprétation construit à partir des formes lexicales qui, rétroactivement, se voient investies du rôle de représenter ce contenu.

C'est sur ce dernier point que le modèle de fonctionnement du lexique en indexation articule les deux représentations lexicaliste et objectiviste : comme toute la langue se comprend à travers ses formes lexicales, tout ce qu'on suppose à son sujet se trouve investi au niveau des mots.

(ii) Les limites du modèle objectiviste en indexation

Après d'autres⁴², on peut relever que les limites imputées à la langue dans le discours classique relèvent des limites propres aux présupposés du modèle objectiviste :

- la langue comme code : le fait que la langue soit appréhendée comme un code conduit les professionnels de l'indexation à dire qu'elle est un « mauvais » code (trop ambigu, trop irrégulier, etc.) mais qu'il est possible d'améliorer ce code moyennant quelques aménagements (réduction du nombre de formes lexicales, explicitation des relations sémantiques entre ces formes, etc., dans les langages documentaires) ;
- la langue comme véhicule d'information pré-existante : ce présupposé invite les professionnels de l'information à faire preuve de grande sévérité envers ce qu'ils nomment la variabilité de l'indexation⁴³. Non seulement on déplore que deux indexeurs ne puissent indexer de façon identique un même document, mais on s'étonne aussi que le même indexeur indexe différemment le même document en T₁ et en T₂. C'est là un problème que la pratique a du mal à résoudre. Comme celui de l'évolution des connaissances : la pratique de l'indexation est prise dans une éternelle spirale du

⁴². Par exemple, Turner 1990, Dubois 1995.

⁴³. Cf. par exemple, Bertrand-Gastaldi 1986, p. 11-13.

« retard » qui rend obsolètes, à peine achevés, les langages documentaires qu'elle met au point. Là encore, c'est à la vision d'une stabilité des connaissances et de l'information, propre au modèle objectiviste, que se heurte la pratique de l'indexation.

Certains voient, dans l'adoption implicite du modèle objectiviste du langage, des contradictions encore plus importantes, touchant la raison d'être même de l'indexation comme moyen de mise à disposition des connaissances ; ainsi Turner se montre-t-il particulièrement virulent :

L'ambiguïté est inhérente au langage. Elle est en grande partie responsable du renouvellement constant des problématiques scientifiques. Vouloir figer la signification des mots par l'adoption d'un symbolisme rigoureux est un non-sens. Un tel projet ne tient aucun compte de la dynamique de la construction sociale des connaissances.
Turner, 1990, [p. 2].

On trouve une semblable critique chez d'autres auteurs, notamment Dubois⁴⁴. Là encore quelle conclusion doit-on tirer de l'adoption implicite du modèle objectiviste dans l'approche du lexique en indexation ?

Il nous semble que l'on doit tenter de s'affranchir de ce modèle, non seulement parce que, n'étant pas valide d'un point de vue théorique, il ne permet pas de dégager des fondements de l'indexation, mais aussi parce que les représentations linguistiques qu'il véhicule enferment la pratique de l'indexation dans de faux problèmes⁴⁵ et l'empêchent de penser ses propres objets : document, information, descripteur, etc.

En outre, concernant précisément le descripteur, le modèle objectiviste s'avère insuffisant pour expliquer certaines de ses caractéristiques, et notamment les caractéristiques 4 et 5 :

- son autonomie : sur quoi repose, dans le modèle objectiviste, le fait que le descripteur puisse être utilisé seul, détaché du document auquel il renvoie ? Sur une autonomie conceptuelle ?
- son rôle de relais textuel : comment le descripteur, expression univoque d'un concept stable, pourrait-il renvoyer à différents concepts issus de plusieurs documents ? Si le descripteur ne doit

44. Dubois 1995, p. 92 : « Elle [l'ambiguïté] est simplement normale, voire même « naturelle », dans la nature des langues comme de toutes les productions humaines [...] elle est en elle-même productive des connaissances. Tenter de réduire systématiquement ces écarts, plutôt que de les repérer et de les gérer comme tels, ferme en effet la possibilité d'évolution tant individuelle que collective des connaissances ».

45. Notamment celui des langages documentaires : les langages sont toujours imparfaits et incomplets, mais, tant que l'on ne disposera pas de langages parfaits et complets, l'indexation restera d'application délicate.

exprimer qu'un seul et unique concept, toujours le même, comment peut-il, en tant que terme préférentiel, rendre compte aussi des concepts de ses quasi-synonymes ?

I.3- Conclusion et résultats intermédiaires

Dans notre étude du modèle d'utilisation du lexique en indexation (I.1), nous avons dégagé, à partir des deux fonctions de l'indexation, cinq caractéristiques :

- **fonction 1 de l'indexation** : « indiquer brièvement, sous forme concise, la teneur du document »
fonction 1 du descripteur : le descripteur comme unité de représentation du contenu d'un document :
 - > caractéristique.1 : le descripteur comme expression linguistique d'un concept pré-existant, stable, unitaire (« simple ») ;
 - > caractéristique.2 : le descripteur comme expression linguistique stable ;
 - > caractéristique.3 : le descripteur comme condensateur textuel.
- **fonction 2 de l'indexation** : « permettre la recherche efficace des informations contenues dans un fonds documentaire »
fonction 2 du descripteur : le descripteur comme accès à un fonds documentaire :
 - > caractéristique.4 : le descripteur comme expression linguistique autonome ;
 - > caractéristique.5 : le descripteur comme relais textuel.

Les caractéristiques 1, 2, 3 du descripteur qui relèvent de la fonction 1 de l'indexation (représentation du contenu d'un document) nous sont apparues très déterminées par l'adoption implicite du modèle objectiviste du langage. Nous avons fait l'hypothèse que ces trois caractéristiques du descripteur n'étaient propres qu'au modèle objectiviste et qu'elles n'étaient plus pertinentes dans un autre cadre. En effet, la fonction dont elle relève - celle de représenter le contenu d'un document - peut être analysée en termes d'effet d'interprétation qui, construit à partir des formes lexicales, se trouve *a posteriori* attaché aux unités mêmes, aux descripteurs. Nous montrons ci-après, dans le point II.1, comment on peut envisager la construction d'un tel effet.

L'indexation est donc, à ce stade de notre recherche, vidée de la notion d'analyse et de représentation du contenu et se voit recentrée autour de sa fonction d'accès à l'information. Sur ce dernier point, on a pu remarquer que les caractéristiques 4 et 5 du descripteur, qui relèvent de la fonction 2 de l'indexation, ne peuvent être traitées de façon adéquate dans le cadre du modèle objectiviste.

Dans le modèle de fonctionnement du lexique que nous proposons ci-après, et dans lequel nous mènerons notre étude de la représentation du contenu comme effet, les caractéristiques 4 et 5 du descripteur pourront être analysées et proposées comme fondements théoriques.

II- Déplacement du modèle de fonctionnement du lexique

Dans ce paragraphe, nous étudierons les cinq caractéristiques du descripteur précédemment dégagées au regard de deux modèles de description linguistique des unités lexicales. Notre objectif est de distinguer, parmi les caractéristiques du descripteur, celles qui relèvent d'effets propres à une utilisation de la langue et celles qui relèvent de principes propres à la langue.

Compte tenu des deux plans mis en jeu dans l'indexation (plan des textes et plan des mots), nous travaillerons à partir des deux modèles linguistiques suivants :

- pour comprendre le rapport que l'indexation établit entre texte et mot, nous étudierons ce que le **modèle de l'analyse de discours** dit du rapport entre thème et discours ; c'est sous cet angle que nous essayerons de distinguer faits et effets dans les caractéristiques 1, 2, 3 et 5 du descripteur (II.1) ;
- pour approcher le « saut » qu'opère l'indexation, par le biais du mot, entre « contenu » d'un document et « information » d'une collection documentaire, nous examinerons le statut des unités lexicales dans le cadre d'une **théorie du lexique*** : cette approche linguistique des unités lexicales hors emploi nous permettra de préciser les caractéristiques 4 et 5⁴⁶ du descripteur (II.2).

Le déplacement de modèle de fonctionnement que l'on propose ici se marque par l'introduction, dans ce paragraphe, de nouvelles notions, proprement linguistiques⁴⁷, pour lesquelles le glossaire situé en fin d'étude propose une première approche.

II.1- Distinction des faits et des effets par le biais du modèle de l'analyse de discours

Sous une forme très simplifiée, nous exposons dans ce paragraphe quelques éléments issus du modèle linguistique de la « nouvelle analyse de discours »⁴⁸. Ce modèle, originellement proposé par Pêcheux dans les années 1970 en réponse aux

⁴⁶. La caractéristique 5 du descripteur (le descripteur comme terme textuel) est abordée sous deux angles : sous l'angle du texte (§ II.1) et sous l'angle du terme (§ II.2).

⁴⁷. On verra apparaître les notions de : unité lexicale, signification lexicale, interprétation d'un texte, thème de discours, objet de discours, référent.

⁴⁸. Abrégée dans ce document sous la forme « analyse de discours ». La nouveauté de l'approche est exprimée notamment dans Marandin 1979 et 1993 et évaluée dans Marandin 1997.

insuffisances des approches alors dominantes⁴⁹, a été particulièrement travaillé par Marandin.

L'approche de Marandin permet de montrer comment se construisent entre mots et textes des effets d'interprétation : des effets de « cohérence thématique » quand un mot est identifié comme thème dans un discours et des effets d'« intertextualité »⁵⁰ quand le mot d'un discours semble appeler ou rappeler d'autres discours. Marandin montre en effet que le thème de discours se construit nécessairement au travers d'autres discours, la notion de discours étant, dans ce cadre, un « construit » : « elle ne renvoie pas à la simple donnée d'un enchaînement d'énoncés »⁵¹.

C'est pour ces raisons que le cadre de l'analyse de discours nous paraît pertinent :

- il dégage une voie pour penser le lien entre les deux fonctions de l'indexation distinguées dans la norme : l'analyse d'un document suppose-t-elle toujours la prise en compte d'autres documents ? Dans ce cadre, comment se pense le lien entre l'analyse d'un document et l'analyse de plusieurs documents ? ;
- il fournit des pistes pour redéfinir l'indexation : comment se traduit le changement de problématique, de l'« analyse de contenu » à la constitution de l'intertextualité ?

II.1.1- La notion de représentation du contenu d'un document : un effet d'interprétation

En disant que la représentation du contenu relève, en indexation, d'un effet d'interprétation, nous ne voulons pas dire que l'indexation ne permet pas, ou permet faussement, de se faire une idée du contenu d'un document. Nous tenons que l'indexation d'un document (en l'occurrence les descripteurs attribués à un document) permet bel et bien de se faire une idée du document mais nous disons que « cette idée » n'est pas liée à l'« idée », au « concept » contenu dans le descripteur, dans le mot lui-même. C'est par le biais de son interprétation en discours que le descripteur permet de rendre compte du contenu d'un document.

Pour expliquer comment des mots (les descripteurs) peuvent donner l'impression, à juste titre sans aucun doute, qu'ils disent le contenu d'un texte, nous présentons

⁴⁹. Notamment celles proposées par les grammaires de textes (qui analysent le texte sur le modèle d'une analyse de la phrase), dont on peut trouver une critique dans Marandin 1979.

⁵⁰. Marandin [1993] rejette explicitement le terme en raison de l'utilisation qui en est faite dans un autre cadre que le sien (cadre structuraliste) ; il propose la notion d'interdiscours, qui n'est cependant pas directement concurrente, voir ci-après, § II.1.2. Nous ne pouvons entrer dans le détail de ces discussions.

⁵¹. Marandin 1997, p. 12.

une recherche menée par Marandin⁵² dans le cadre de l'analyse de discours (A). Nous aborderons ensuite (B) la question de savoir si, en indexation, un descripteur peut être pour un document ce qu'un thème est, en analyse de discours, pour un discours.

A- Le thème de discours dans le cadre de l'analyse de discours

Marandin [1988] s'intéresse à la notion de thème de discours sous l'angle de la problématique suivante :

- qu'est-ce qu'un thème de discours ?
- comment se matérialise-t-il dans le discours ?
- comment intervient-il dans la compréhension d'un texte ?

Selon lui, le point de vue sur le thème est nécessairement celui de l'interprétation (point de vue de la réception) ; la question centrale peut se résumer alors de la façon suivante : comment détermine-t-on un thème de discours ?

Ce qui nous intéresse dans la démarche de Marandin, c'est qu'elle rend compte de la formulation lexicale du thème dans des termes autres que ceux de la réduction, de la condensation d'un texte par des mots. Parallèlement, c'est une approche qui dit qu'un mot peut rendre compte d'un texte, et même que l'on ne peut rendre compte d'un texte que par des mots, qu'en le nommant et le renommant⁵³. C'est ainsi que cette approche peut permettre de comprendre comment texte et mot peuvent être assimilés, confondus, et comment l'on peut *a posteriori* désigner le mot comme porteur de l'interprétation d'un texte.

Avant d'en venir à l'explication du mécanisme de l'assimilation thème/discours, nous présentons ci-après les principales propositions de Marandin :

- (i) un thème, c'est toujours un **nom**, et plus précisément, un individu linguistique de type syntagme nominal (abrégé SN ci-après, appelé aussi GN, groupe nominal, par Marandin) ;
- (ii) un thème, c'est un nom qui a la **capacité d'organiser** un texte : cette approche du thème s'oppose à celle qui voit dans le thème un nom

⁵². Marandin 1988, p. 67-87.

⁵³. Marandin cite à ce sujet Barthes : « Quiconque lit un texte rassemble certaines informations sous quelques noms génériques et c'est ce nom qui fait la séquence ; la séquence n'existe qu'au moment où et parce qu'on peut la nommer, elle se développe au rythme de la nomination qui se cherche et se confirme », Barthes 1970, p. 14 (cité in Marandin 1997, p. 21). Dans Marandin 1988, l'article s'ouvre sur cette autre citation, proche de la précédente dans l'esprit, de Barthes : « Lire, c'est trouver des sens, c'est les nommer ; mais ces sens nommés sont emportés vers d'autres noms ; les noms s'appellent, se rassemblent et leur groupement veut de nouveau se faire nommer : je nomme, je renomme ; ainsi passe le texte : c'est une nomination en devenir, une approximation inlassable, un travail métonymique », Barthes 1970, p. 17 (cité in Marandin 1988, p. 67).

qui dit ce « à propos de quoi » est le texte⁵⁴ ;

(iii) organiser un texte, c'est en donner une interprétation⁵⁵ : le thème est à ce titre un SN apte à **projeter une interprétation** ; on remarque là encore que, dans cette approche, le thème n'est pas un nom qui interprète un texte ou une partie de texte.

La forme lexicale d'un thème de discours n'est donc pas une interprétation de ce discours mais plutôt un élément déclencheur de la construction d'une interprétation. Thématiser revient alors à constituer des unités d'interprétation et non à retrouver dans un texte des unités déjà interprétées (les éléments d'une encyclopédie dans les termes de Marandin⁵⁶).

C'est à ce titre que le thème est nécessairement tissé à même la trame du texte : c'est toujours un terme textuel, un terme par rapport à un texte, que le nom du thème soit issu du texte ou qu'il soit déterminé de l'extérieur, dans l'espace de réception du texte⁵⁷. Dans les deux cas, c'est le texte qui dispose de quoi constituer le thème.

C'est cette caractéristique du thème comme terme textuel qui est à l'origine de l'« illusion d'optique », de l'effet d'eschérisation dit encore Marandin, que génère le thème :

La possibilité pour un groupe nominal de désigner un individu dans un monde, quand il est traité ou interprété dans un énoncé d'occurrence, et le monde dans lequel s'identifie cet individu, quand il est traité comme thème de discours.
Marandin, 1988, p. 82.

C'est ce mécanisme d'identification entre individu et monde, c'est-à-dire entre thème et discours, que nous allons essayer d'expliquer en reprenant l'analyse et les exemples de Marandin.

Le point de l'analyse porte sur l'approche du terme textuel comme un terme par rapport à un texte. Ce qui fait un terme textuel, c'est le « contenu descriptif » d'un terme. Un contenu descriptif est relatif à un discours donné : il correspond à ce qui est introduit par les énoncés et ensuite transformé en propriétés du terme.

54. Voir Marandin 1997, p. 21 : « Le thème de discours est un aspect du processus de compréhension et non pas la donnée d'un individu externe à propos de quoi le discours se tient ».

55. Un thème est en cela une manière de comprendre. Il y a toujours plusieurs manières de comprendre et donc plusieurs thèmes possibles ; la question reste de « déterminer comment une même suite d'énoncés permet des lectures différentes et ce qui, dans les énoncés ou leurs enchaînements, oriente vers telle lecture plutôt que vers telle autre ». Marandin propose des pistes, 1988, p. 71 et suiv.

56. *Ibid.*, p. 84.

57. C'est la différence entre thème configuré et thème inféré qu'établit Marandin 1988 p. 77 et suiv. Nous l'évoquons ci-après.

Soit l'exemple suivant repris de Marandin 1988 :

La licorne à fourrure d'hermine abondait autour du château. Un jour, Lancelot s'amusa à les pourchasser. Piqué au jeu, il les tua toutes. Puis, il les dépouilla et il s'empara de leur précieuse toison. Trois jours après, il mourrait dans d'affreuses douleurs. Lancelot fut pleuré de tous. Isolde s'enferma dans un couvent...

Dans cet exemple, on peut dégager deux types de thème : un thème configuré (nommé dans le texte) et un thème inféré (nommé dans l'espace de réception du texte).

Comme thème configuré, il y a par exemple *Lancelot*, dont le contenu descriptif est /celui qui s'appelle Lancelot, qui a pourchassé, tué, dépouillé les licornes et qui est mort après/. Si *Lancelot* est le thème de ce discours, on dira que ce discours n'est pas « à propos de » Lancelot mais « à propos de » /celui qui s'appelle Lancelot, qui a pourchassé, tué, dépouillé les licornes et qui est mort après/, etc.

Comme thème inféré, il y a par exemple *la vengeance des licornes*, dont le contenu descriptif, « abstrait de la suite d'énoncés »⁵⁸, peut être /la mort de Lancelot est liée à ce qu'il a fait trois jours auparavant, les licornes ont causé la mort de Lancelot, les licornes font mourir Lancelot pour venger un méfait dont elles sont les victimes, leur extermination étant un méfait à leur rencontre/. Là encore, si *la vengeance des licornes* est un thème du discours, on dira que ce discours n'est pas « à propos de » la vengeance des licornes mais « à propos de » /la mort de Lancelot est liée à ce qu'il a fait trois jours auparavant, les licornes ont causé la mort de Lancelot, les licornes font mourir Lancelot pour venger un méfait dont elles sont les victimes, leur extermination étant un méfait à leur rencontre/.

Le contenu descriptif du terme qui en fait un terme textuel et par suite un thème de discours peut donc être défini comme « un agrégat subsumant d'autres individus dans leurs interrelations, telles qu'elles sont **introduites dans les énoncés, reconstruites dans la compréhension et constitutives d'une interprétation** »⁵⁹.

Cette approche du thème de discours comme terme textuel nous paraît précieuse en ce qu'elle fait apparaître deux dimensions de la thématization :

- une dimension de l'interprétation⁶⁰, qui affaiblit l'idée qu'un texte aurait, hors lecture, en soi, un thème, des thèmes et que, hors

⁵⁸. Marandin 1997 [p. 23] met l'accent sur ce qui permet dans un texte d'inférer un thème : ce sont les *transitions temporelles* ; pour une description, voir Marandin 1988, p. 81.

⁵⁹. Marandin 1988, p. 82.

⁶⁰. Nous n'avons pas vraiment développé dans ce paragraphe cet aspect de la recherche de Marandin [1988, p. 70], c'est-à-dire son idée que « la problématique du thème de discours recoupe celle de l'anaphore comprise au sens de " dépendance interprétative " que contractent entre eux les GN d'un texte ». Elle repose sur la notion de chaîne de référence, que nous étudierons au chapitre V.

- lecture, en eux-mêmes, ces thèmes pourraient dire un contenu ;
- une dimension de la transformation : la transformation d'énoncés en propriétés permet de comprendre comment un terme isolé de son contexte peut donner l'impression de dire le contenu d'un texte, ou d'avoir un contenu tout court.

Cette analyse du thème de discours qui démonte le mécanisme de l'assimilation thème/discours peut-elle être reprise dans le cadre de l'indexation ?

B- Le thème de discours en indexation : première approche⁶¹

Pour que la recherche de Marandin prenne un sens dans notre cadre, il faut que le descripteur puisse être considéré comme un thème de discours.

C'est là une proposition qui n'est pas nouvelle en indexation. En effet, Michel Le Guern⁶² a pu défendre que :

- le descripteur est une unité de discours ;
- le descripteur est, plus précisément, un syntagme nominal ;
- l'indexation peut être vue comme la détermination de thèmes⁶³.

L'hypothèse du descripteur comme thème de discours est donc déjà posée ; nous la reprenons à notre compte dans cette recherche et nous donnons pour objectif de la préciser, notamment en examinant :

- comment un document peut être considéré comme un discours, objet du chapitre IV ;
- comment (et par qui) le descripteur se constitue au fil du discours, à même le texte, comme un terme textuel, objet du chapitre V.

Reste qu'adopter l'hypothèse du descripteur comme thème de discours n'est pas, même dans le cadre d'un modèle d'utilisation (d'un modèle qui utilise le concept linguistique de thème), sans poser des problèmes, notamment de représentativité du phénomène : nous en abordons deux ci-après.

(i) Tous les descripteurs peuvent-ils être perçus comme des thèmes de discours ?

À simplement regarder une formule d'indexation telle que la produisent les organismes documentaires, on remarque que tous les descripteurs utilisés ne peuvent fonctionner comme thèmes de discours. Ainsi dans cette indexation⁶⁴ réalisée par la *Documentation française* à partir d'un article du *Monde*⁶⁵ :

⁶¹. Une approche plus approfondie est proposée dans le chapitre V consacré au descripteur.

⁶². Le Guern 1984, notamment.

⁶³. *Ibid.*, p. 168.

⁶⁴. Voir annexe 2 : les données de notre enquête.

⁶⁵. *Le Monde* du 1/12/1994, p. 24.

Tapie Bernard, Bernard Tapie Finances, Groupe Bernard Tapie, tribunal de commerce, jugement, redressement judiciaire, délai, personnalité position, Crédit Lyonnais, affaire, navire, fortune, cour d'appel, Testut, amende, sanction, COB.

On remarque que - et c'est là une particularité des langages documentaires qui tend à opacifier le mécanisme de l'indexation quand elle y a recours - tous les descripteurs utilisés en indexation ne jouent pas le même rôle par rapport au document : tous ne décrivent pas le « contenu », tous ne sont pas des « descripteurs matières ». C'est le cas de ceux que l'on appelle les « mots-outils »⁶⁶ (dans l'exemple : « affaire » et « délai ») ou encore les « mots-facettes »⁶⁷ (dans l'exemple : « personnalité position »).

L'indexation effectuée avec le langage documentaire Rameau⁶⁸ est sur ce point tout à fait exemplaire. Une « vedette-matière » (la formule d'indexation complète) est constituée systématiquement d'éléments de nature hétérogène :

- une « tête de vedette » qui correspond à un mot (ou plusieurs) qui « exprime » le « sujet » d'un document : par exemple « cerveau » ;
- des subdivisions, qui sont des mots chargés de préciser la tête de vedette sous différents points de vue : point de vue du sujet (par exemple « maladies » pour le sujet « cerveau ») ; point de vue géographique (par exemple « Italie ») ; point de vue chronologique (par exemple « Renaissance ») ; point de vue de la forme du document (par exemple « répertoires »).

Ainsi, dans la formule d'indexation suivante issue de Rameau - « cerveau**maladies**Italie**Renaissance**répertoires »⁶⁹ -, chacun des mots est à lire et à interpréter d'une façon particulière, qui, certes, est spécifiée dans les gros volumes⁷⁰ qui accompagnent le répertoire du vocabulaire Rameau, mais qui ne présente aucun caractère d'évidence.

Il y a sans aucun doute, sur le point précis des langages documentaires, toute une série d'analyses de détail à mener qui permettraient de mettre au jour d'autres aspects de la perception du lexique chez les professionnels de la documentation.

⁶⁶. « Descripteur n'ayant pas de valeur documentaire spécifique et généralement utilisé en association avec un ou plusieurs descripteurs », norme Z 47-100 (1981), p. 19.

⁶⁷. « Catégories de notions de même nature tels que processus, phénomène, matériau, outil, permettant un regroupement de termes indépendamment des disciplines traitées ». *Id.*

⁶⁸. Répertoire d'autorité matière encyclopédique et alphabétique.

⁶⁹. Pour des raisons peu claires, le langage Rameau préconise l'usage du pluriel, « sauf quand l'usage l'interdit » ; les exemples donnés sont tous des noms propres : « jugement dernier », « communion des saints », « France ».

⁷⁰. Guide d'indexation RAMEAU [Bibliothèque nationale de France 1995].

Pour ce qui concerne notre sujet de recherche - l'indexation -, il nous paraît essentiel de pouvoir prendre des distances par rapport aux présupposés propres à tel ou tel langage documentaire : nous considérons la notion de langage documentaire dans son principe et, à ce titre, nous constatons qu'elle constitue une difficulté pour concevoir le descripteur de façon unifiée (par la notion de thème de discours). Mais cette difficulté ne constitue pas pour autant un obstacle, notamment si on peut montrer, ne serait-ce que partiellement, que l'indexation peut réaliser son objectif en se passant de l'usage d'un langage documentaire. L'importation de la notion de thème de discours dans un modèle d'utilisation du lexique en indexation n'est donc pas nécessairement entravée par les modes de fonctionnement particuliers que mettent en œuvre les langages documentaires. Si elle est problématique, c'est essentiellement dans un cadre où l'indexation ne se pense qu'au travers d'un langage documentaire ; or cette position empêche de définir en propre le descripteur qui - on l'a vu - se donne sous une forme singulièrement hétérogène.

(ii) Tous les aspects du descripteur peuvent-ils être rendus par le notion de thème de discours ?

Un autre type de problème peut faire apparaître l'importation de la notion de thème de discours en indexation comme sévèrement restrictive. En effet, l'indexation doit, comme nous l'avons vu, permettre d'accéder à l'information d'un fonds documentaire et, à ce titre, le descripteur doit pouvoir être appréhendé à la fois comme une unité autonome et comme un synonyme (fonction 2 de l'indexation) : nous aurons donc à réinterroger la notion de thème de discours sous l'angle de ces deux aspects du descripteur⁷¹.

Ce paragraphe (II.1.1) avait pour objet de préciser comment s'établit la notion de représentation de contenu en indexation.

L'analyse du thème de discours proposée par Marandin montre comment un thème peut représenter un discours : par le biais d'une illusion d'optique déclenchée sur la base du contenu descriptif d'un terme compris comme terme textuel. Si l'on adopte l'hypothèse que le descripteur peut se comprendre, dans le cadre d'un modèle d'utilisation du lexique, comme un thème de discours, alors on peut poser que l'analyse de contenu en indexation relève de cette même illusion d'optique, d'un effet déclenché sur la base des descripteurs par le biais d'une interprétation : nous y reviendrons.

⁷¹. Voir, sur ce point, le chapitre V.

Cette approche nous conduit à voir que la notion de représentation de contenu en indexation, si elle relève de l'interprétation (donc du niveau de la réception), ne peut être donnée pour un principe de production : si les descripteurs en tant que termes textuels donnent la possibilité de représenter le contenu, ils ne sont pas eux-mêmes des unités de représentation de contenu.

Cherchant à montrer que la représentation du contenu relevait en indexation d'un effet d'interprétation, nous avons du même coup posé une nouvelle hypothèse et ouvert une nouvelle piste de recherche :

- l'hypothèse du descripteur comme thème de discours, hypothèse qui nous éloigne de la conception objectiviste d'une transmission stabilisée de concept, de l'auteur à l'utilisateur : il apparaît en effet que le descripteur intéresse moins pour son « contenu » propre (conceptuel ? sémantique ?) que pour le contenu textuel (le « contenu descriptif ») auquel il permet d'accéder ;
- cette hypothèse engage la recherche dans des voies nouvelles : qu'est-ce que l'indexation met à disposition de l'utilisateur pour qu'il construise son interprétation ? En quoi les accès aux documents qu'elle dispose sont-ils à même de permettre la construction de ces interprétations ? L'approche du descripteur comme thème de discours est-elle suffisante pour caractériser l'ensemble du fonctionnement attendu du descripteur ?

C'est à travers les données de l'analyse de discours que nous avons proposé de comprendre la notion de représentation de contenu en indexation comme un effet, un effet d'interprétation qui se donne, dans le discours classique, pour un principe de production.

C'est à nouveau par l'analyse de discours que nous examinerons comment le descripteur peut fonctionner comme un relais textuel, c'est-à-dire comment un mot peut rapprocher plusieurs textes différents.

II.1.2- La notion d'accès à un fonds documentaire : un principe d'interprétation en analyse de discours

Nous avons précédemment dégagé du modèle d'utilisation du lexique en indexation cinq caractéristiques du descripteur, réparties en deux groupes correspondant aux deux principales fonctions de l'indexation. Parmi ces caractéristiques, se trouve celle du descripteur comme relais textuel, c'est-à-dire comme forme lexicale susceptible de lier entre eux plusieurs documents différents. Nous avons noté que

le modèle objectiviste du langage se révélait insuffisant pour décrire à la fois la possibilité et le fonctionnement de cet aspect du descripteur. Le cadre d'analyse que nous avons précédemment adopté, celui de l'analyse de discours, nous permet de traiter cette caractéristique du descripteur, et donc de traiter l'ensemble les deux fonctions de l'indexation dans les mêmes termes : c'est sur cette base que nous pourrions avancer une hypothèse qui explique le lien entre les deux aspects de l'indexation⁷².

Si le paragraphe précédent traitait de l'écart qui maintient distincts thème/discours et des effets interprétatifs qui peuvent les lier, celui-ci traite de l'écart qui maintient distincts discours₁/discours₂...discours_n et des effets interprétatifs qui, toujours à partir d'un mot, peuvent les mettre en relation. Si les problématiques ne sont pas exactement les mêmes, elles présentent des points communs. Les deux reposent sur l'appréhension des formes lexicales comme « relais », mais à deux niveaux différents :

- la relation mot/texte engage la problématique à un niveau de l'intradiscours ;
- la relation texte/texte par le mot l'engage à un niveau de l'interdiscours.

A- Les notions d'interdiscours et intradiscours en analyse du discours

L'un des traits majeurs du courant de l'analyse de discours dans lequel s'inscrit Marandin consiste à établir une relation fondamentale entre deux niveaux : l'intradiscours, l'analyse d'un discours, ne tient que par l'interdiscours, l'analyse à travers plusieurs discours. L'enjeu peut être ainsi formulé :

On peut bien dire que l'intradiscours en tant que le "fil du discours" du sujet est strictement un effet de l'interdiscours sur lui-même, une "intérieurité" entièrement déterminée comme telle de l'"extérieur".
Pêcheux 1975, cité in Marandin 1997, p. 12. C'est nous qui soulignons.

La présence d'autres discours dans un discours se marque par les mots, qui n'entrent jamais « seuls » dans les énoncés, mais qui ouvrent au contraire « la mémoire et l'anticipation d'autres textes »⁷³. Le thème discursif, se construit au fil du texte mais dans un texte, il y a toujours des « hiatus sémantiques », des espaces où peuvent se glisser des bribes d'autres textes.

Dans ce type d'analyse de discours, un mot n'est jamais, à proprement parler, mis face à un texte. L'analyse du thème de discours ne relève en effet pas d'une

⁷². On se souvient que, dans la norme, le lien entre les deux fonctions n'est pas spécifié (*supra*).

⁷³. Marandin 1984, p. 53.

sémantique lexicale, mais d'une sémantique discursive⁷⁴ :

On retrouve ici le point de départ historique de l'analyse de discours : une réflexion sur le mot dans la lexicographie politique, puis dans la critique de la sémantique lexicale effectuée par Pêcheux dans Les Vérités de la Palice. Mais la problématique s'est déplacée d'une approche sémiotique où le mot est élément de système (de langue ou idéologique) à une problématique sémantique où il est marqué (chargé dirait Bakhtine) par ses occurrences dans d'autres textes (les usages qui en sont faits ou les textes qui s'y condensent).

Marandin, 1984, p. 53.

Si toute analyse de discours suppose l'analyse de plusieurs discours, les mots du discours analysé sont nécessairement analysés aussi dans d'autres discours. Dès lors qu'un même mot, qu'un même thème de discours, puisse renvoyer à des discours différents ne surprend plus puisqu'il n'a été constitué comme tel qu'à partir d'une approche interdiscursive.

Si donc le descripteur peut fonctionner, dans un cadre qui reste à définir, comme thème de discours, alors il fonctionne nécessairement comme relais textuel : ce n'est là qu'un autre effet, quoiqu'à un autre niveau, du fait qu'il soit thème de discours.

B- Les notions d'interdiscours et d'intradiscours en indexation : première approche⁷⁵

On voit donc que les caractéristiques du descripteur, si on le considère comme thème de discours, sont moins divergentes qu'il n'y paraît. Que le descripteur puisse être considéré comme représentant le contenu d'un texte (fonction 1 de l'indexation) comme celui de plusieurs textes qu'il met en relation (fonction 2) relève pareillement d'effets interprétatifs, de nature différente, mais réalisés sur la même base : celle de la construction du thème de discours, compris dans un cadre où l'intradiscours suppose l'interdiscours. Du coup, la question qui concerne en propre l'indexation devient celle de la production, de la construction de ces effets d'interprétation : comment l'indexation permet-elle de créer des thèmes de discours qui se définissent par leur cheminement à travers plusieurs discours ?

On voit que, si le descripteur est un thème de discours, l'ensemble des discours (des textes, des documents) qui permettent de le constituer comme tel prend une importance déterminante : à ce titre, l'indexation pourrait être comprise comme une construction de l'interdiscours⁷⁶.

⁷⁴. C'est exactement la position que défend Le Guern [1984, 1991a par exemple] pour l'approche du descripteur en indexation : nous y revenons en III.

⁷⁵. Une approche plus approfondie est proposée dans le chapitre IV.

⁷⁶. C'est une hypothèse que nous défendons dans le chapitre IV.

En effet, il est clair qu'un thème de discours ne se construit pas à travers n'importe quels discours ; il est nécessaire de constituer un ensemble de discours susceptibles de s'entrecroiser :

On peut se donner pour objectif de suivre la trace [des] textes présents, absents dans le texte de départ. Il faut alors définir une **stratégie d'exploration**. Diverses stratégies sont concevables. L'une d'elle se définit sur le modèle co-textuel (...) : des fragments de textes sont rassemblés parce qu'ils se donnent dans les formes que l'analyse intra-textuelle décrit comme semblables.
Marandin, 1984, p. 54. C'est nous qui soulignons.

Si le descripteur est un thème de discours, un thème issu de l'interdiscours, l'indexation porte donc non plus sur un document mais sur plusieurs documents, un fonds documentaire ; et l'essentiel de sa stratégie consiste moins à déterminer le contenu d'un document qu'à déterminer des principes de regroupement des documents. À la suite de Marandin en analyse de discours, nous proposerons des principes de regroupement qui empruntent leurs formes aux « formations discursives » proposées par Foucault (*infra* chapitre IV).

II.1.3- Conclusion et résultats intermédiaires

Les questions qui se posaient au début du paragraphe II de ce chapitre étaient les suivantes : les caractéristiques du descripteur dégagées de notre lecture de la norme sont-elles définitoires du descripteur ? Peut-on à partir d'elles constituer des fondements théoriques de l'indexation ? Que deviennent, dans le cadre d'une approche linguistique du lexique, les caractéristiques du descripteur qui semblaient si fortement déterminées par le modèle objectiviste sous-jacent au discours classique ?

Le paragraphe II.1 nous a permis de répondre à la dernière de ces trois questions et partiellement aux deux premières. Nous avons proposé de considérer les caractéristiques 1, 2 et 3 du descripteur, et plus globalement la fonction 1 de l'indexation, comme relevant d'un principe d'interprétation, considéré *a posteriori* comme un principe de production dans le discours classique : ce n'est qu'*a posteriori* et sur la base des présupposés lexicaliste et objectiviste que l'on peut attribuer aux seuls mots l'essentiel du processus de l'indexation. De la même façon, la caractéristique 5 du descripteur apparaît, à la lumière des propositions de l'analyse de discours, comme un effet d'interprétation. À ce titre, ces caractéristiques ne constituent pas des propriétés définitoires du descripteur et ne peuvent permettre d'établir les principes de l'indexation.

Dans le paragraphe suivant, nous complétons notre réponse à nos deux premières questions en étudiant la fonction 2 de l'indexation (fournir des accès au document, à plusieurs documents).

II.2- Propriétés remarquables des unités lexicales

Notre lecture de la norme nous a permis de dégager deux caractéristiques du descripteur pour que l'indexation puisse réaliser sa fonction de fourniture d'accès à un document comme à plusieurs documents :

- Le descripteur doit révéler une certaine autonomie pour fonctionner seul, être détaché de son contexte : quel type d'autonomie est ici en jeu ? Cet aspect reste implicite dans la norme : il est donné comme une évidence. Le modèle objectiviste suggère que l'autonomie du descripteur pourrait être de nature conceptuelle. Or nous avons dû mettre de côté la notion de « concept », qui nous est apparue trop floue car très large : pouvons-nous penser l'autonomie du descripteur dans d'autres termes ? Pourquoi cette autonomie peut-elle être perçue de façon si évidente ?
- Le descripteur doit permettre de créer des classes d'équivalence. La norme ne dit pas clairement de quel type d'équivalence il s'agit : la notion d'équivalence lexicale semble être mise de côté, mais la notion d'équivalence documentaire, entre des documents, n'est pas explicite. Sur ce point, le modèle objectiviste, sous-jacent au discours de la norme, ne nous aide guère. Il pointe au contraire les limites d'une approche conceptuelle du descripteur : comment une expression censée représenter un seul et unique concept de façon stable pourrait-elle aussi représenter d'autres concepts, eux-mêmes uniques, de façon tout aussi stable ? En mettant de côté, là encore, la notion de concept, peut-on dégager des modes de fonctionnement du descripteur qui expliquent comment puisse se réaliser cette multi-désignation ?

C'est à partir de ces deux questions que nous examinerons la théorie des unités lexicales hors emploi : deux propriétés remarquables se dégagent, l'une concerne l'autonomie lexicale, l'autre la synonymie.

II.2.1- La possibilité d'utiliser des unités lexicales hors emploi : la question de l'autonomie lexicale

A- Approches de l'autonomie lexicale

La question de l'autonomie des unités lexicales se pose dans le cadre des théories du lexique entendues comme études des unités lexicales hors emploi⁷⁷.

Si la question de l'autonomie lexicale est pertinente dans ce contexte, elle reçoit des modes de description différents selon les approches. On peut distinguer principalement trois positions⁷⁸ :

- (i) l'autonomie d'une unité lexicale est de nature référentielle, cette position est représentée par Kayser 1987 ;
- (ii) l'autonomie d'une unité lexicale est de nature « opératoire », cette position est tenue par Franckel 1992 ;
- (iii) l'autonomie d'une unité lexicale est de nature sémantique, cette position est défendue par Marandin 1992.

Ces trois positions sont en réalité plus intriquées qu'il n'y paraît, les deux premières se constituant sur une critique de la troisième.

Les positions (i) et (ii), qui se différencient par le mode de résolution qu'elles proposent, abordent en effet la question de l'autonomie de l'unité lexicale sous le même angle : comment rendre compte de la multiplicité des objets (*i.e.* référents) auxquels renvoie une même unité lexicale (i) ? Comment rendre compte de la multiplicité d'emplois d'une même unité lexicale (ii) ? Ces deux positions se fondent sur une critique de la signification lexicale qui, pour être communément donnée comme unique et complète, se révèle pourtant inapte à rendre compte de la multiplicité, des référents ou des emplois :

- (i) pour Kayser, la « sémantique n'a pas de sens »⁷⁹ : le « sens unique » que l'on attribue généralement à une unité lexicale (la définition lexicographique par exemple) conduit à une « impasse » en ce qu'il est incapable de rendre compte de la multiplicité des catégories référentielles auxquelles peut renvoyer une unité lexicale ;

⁷⁷. Dans le vaste champ de la recherche linguistique, la question de l'autonomie d'une unité lexicale constitue elle-même une problématique. Nous n'aborderons pas ici les arguments qui vont contre la possibilité d'étudier les unités lexicales hors emploi.

⁷⁸. Pour l'ensemble de ce paragraphe, on s'inspire du compte rendu de deux colloques organisés par le CELEX (Centre d'études sur le lexique / CNRS) ; le premier, tenu en 1990, portait sur la « définition », le second en 1992 sur l'« individualité lexicale ». On s'inspire plus particulièrement de la seconde série de travaux où trois auteurs (Milner, Franckel, Marandin) confrontent leurs positions à partir d'un même projet : la description hors emploi des unités lexicales.

⁷⁹. Kayser 1987.

(ii) pour Franckel, un terme⁸⁰ n'a pas de signification stable : tout terme « est *a priori* susceptible de contribuer à l'émergence d'une multiplicité de valeurs sémantiques qui ne s'engendrent que par interaction avec l'environnement contextuel »⁸¹.

C'est ainsi que ces deux approches proposent des modes de description des unités lexicales qui se passent de la notion de sens :

(i) Kayser propose un modèle qui se fonde sur les interprétations référentielles possibles d'une unité lexicale. Par exemple, pour « livre », on dispose des choix suivants : « lecture de », « écriture de », « reliure de », etc. Les choix peuvent être effectués automatiquement, par un générateur qui explore successivement les nœuds d'un réseau ; à chaque nœud, une interprétation référentielle, dont on vérifie la compatibilité avec le contexte d'occurrence du terme analysé. Dans ce modèle, une interprétation (un nœud) est déclarée satisfaisante si elle ne débouche pas sur une contradiction ;

(ii) Franckel propose un modèle qui se fonde sur des « schèmes opératoires »⁸² : « chaque terme correspond à un schème particulier, c'est-à-dire à une configuration spécifique de paramètres »⁸³. Les paramètres sont principalement de deux ordres : S (sujet énonciateur) et T (espace, temps). Au terme d'une analyse dans laquelle nous n'entrons pas ici⁸⁴, on pourra déterminer le schème opératoire du terme *porter* par exemple comme étant : « la construction de la localisation de X par Y dans le temps relativement à la non-localisation de X par Y hors du plan temporel »⁸⁵.

Il ne nous appartient pas de discuter ces modèles de description non sémantiques des unités lexicales hors emploi. En revanche, on peut relever, avec Marandin⁸⁶, que ces deux modèles posent la question de la signification lexicale dans les termes d'une alternative que l'on peut tenir pour réductrice ; en effet,

80. Dans ce texte, Franckel [1992, p. 18] utilise le mot « terme » pour désigner « de façon délibérément vague toute entité lexico-morphologique ».

81. *Id.*

82. Ce type d'analyse s'inscrit dans le cadre de la théorie de Culioli, qui cherche à « concilier la notion d'invariance que suppose la notion même de propriétés intrinsèques (hors emploi) d'un terme, et celle de déformabilité et d'instabilité sémantiques à laquelle renvoie la multiplicité des valeurs que, par cette interaction, ce terme contribue en règle générale à engendrer ». *Ibid.*, p. 21-22.

83. *Ibid.*, p. 22.

84. On renvoie à Franckel 1992, p. 26-36.

85. *Ibid.*, p. 34.

86. Marandin 1990, p. 290.

- soit les mots changent de signification selon les discours où ils sont employés : on en déduit qu'il n'y a pas de signification lexicale stable, et donc pas de signification lexicale tout court ;
- soit les mots ne changent jamais de signification : mais cette approche ne tient pas au regard des faits.

On peut aussi poser la question de l'autonomisation des unités lexicales par la signification dans un autre cadre. La signification lexicale reste un facteur individuant ; si elle ne peut rendre compte de phénomènes complexes (multiplicité référentielle ou sémantique), c'est plutôt parce qu'elle est elle-même complexe, « trop massive » : « elle recouvre des phénomènes qu'il importe de distinguer »⁸⁷. Autrement dit, si la signification lexicale est considérée comme complexe et hétérogène⁸⁸, elle comprend alors plusieurs dimensions qui peuvent varier différemment et produire des effets divers. Ce n'est pas pour autant que la notion de signification lexicale n'est pas valide pour penser l'autonomie des unités lexicales.

On adoptera dans cette recherche, sans la justifier davantage, la position (iii) : l'autonomie des unités lexicales est de nature sémantique⁸⁹. Autrement dit, on dira que l'autonomie des unités lexicales tient à leur signification lexicale, qui en conséquence doit présenter des caractéristiques de stabilité.

Dans le cadre adopté, l'une des façons d'aborder la notion de signification lexicale est de la penser par la notion de « stéréotype » proposée par Putnam comme étant l'une des dimensions possibles de la signification lexicale. Plus précisément, la notion de « stéréotype » constitue une « hypothèse sur la signification lexicale du point de vue de l'acquisition du langage »⁹⁰. Autrement dit, si elle ne permet pas de répondre directement à la question : « qu'est-ce que la signification ? », elle donne des éléments de réponse à la question : « comment comprend-on les mots ? ». C'est en particulier pour cette raison qu'elle nous paraît pertinente pour notre étude des faits d'indexation.

Après avoir présenté la notion de stéréotype, nous montrerons la façon dont elle peut participer à la signification lexicale dans le cadre d'une approche

⁸⁷. *Id.*

⁸⁸. Marandin 1992a [présentation], p. 9 : « En tant qu'instance d'individuation, la signification lexicale est complexe et hétérogène : elle est externe aux unités lexicales (schèmes) et spécifique à chaque unité (stéréotypes sur les dimensions construites à partir des schèmes) ».

⁸⁹. Ou plutôt l'autonomie des unités lexicales est principalement de nature sémantique. Il y a, en effet, dans le cadre de ce modèle, deux autres facteurs d'individuation des unités lexicales : la forme phonologique et l'appartenance catégorielle (Milner 1989, p. 324) ; mais il reste que « le sens lexical est un facteur de différenciation *sine qua non* des atomes lexicaux », *Ibid.* p. 345.

⁹⁰. Marandin 1990, p. 286.

morphologique qui étudie la construction du sens de certains types de mots : les mots construits.

B- Approche de la notion de stéréotype

Si Putnam [1990a et 1990b] aborde le problème de la signification lexicale par le biais de la problématique « classique » de la stabilité du vocabulaire ordinaire, il oriente la question d'un point de vue particulier : comment la signification d'un mot demeure-t-elle identique alors que changent les croyances, les théories scientifiques, les usages ?

D'emblée est donc posée une distinction entre connaître le mot et connaître le fait que ce mot désigne. La viabilité de cette distinction est fondée sur ce que Putnam appelle la « division linguistique du travail » : l'idée est qu'il n'est pas nécessaire que tous les locuteurs connaissent les différences précises qui distinguent les « mots », « ils peuvent toujours se fier à des experts qui le feront à leur place »⁹¹. En ce sens, le langage est dit « coopératif », par opposition à la vision individualiste du langage proposée par les mentalistes⁹². Ce que souligne la notion de division linguistique du travail, c'est que le fait, la référence ou encore l'extension* n'a pas à faire partie de la signification pour qu'un locuteur puisse employer un mot. Ainsi est posée la possibilité d'une signification lexicale autonome, distincte des référents qu'une unité lexicale peut permettre de désigner :

Si communiquer la signification du mot " tigre " impliquait que l'on communique la totalité de la théorie scientifique acceptée, ou même la totalité de ce que je crois, à propos des tigres, ce serait une tâche impossible. C'est vrai que lorsque je dis à quelqu'un ce qu'est un tigre, " je lui dis simplement certaines phrases ". [...] Le problème est donc bien quelles phrases ?
Putnam, 1990b, p. 299.

Le stéréotype correspond précisément à cet « ensemble de phrases », phrases qui décrivent une « théorie » extrêmement simplifiée à laquelle il n'est pas nécessaire de croire mais dont on doit **savoir** qu'elle est associée à un terme⁹³. En ce sens, le stéréotype est « une idée conventionnelle, qui peut être fausse, sur un segment de la réalité [...] associée à un mot du langage naturel »⁹⁴. Étant une convention, un « ensemble de croyances partagées »⁹⁵, le stéréotype définit moins un objet que la représentation de cet objet pour une communauté linguistique⁹⁶.

91. Putnam 1990a, p. 54.

92. Les mentalistes laissent entendre que « tout ce qui est nécessaire à l'usage du langage est emmagasiné dans chaque esprit individuel », Putnam 1990a, p. 57.

93. Putnam, 1990b, p. 300.

94. Marandin 1990, p. 285.

95. Putnam 1990a, p. 96.

96. Fradin et Marandin 1979, p. 65.

C'est pourquoi le stéréotype peut être faux :

Je peux référer à une espèce naturelle avec un terme qui est " chargé " d'une théorie dont on sait qu'elle n'est plus vraie de cette espèce, car tout le monde sait bien que mon intention est de référer à *l'espèce en question* et non de soutenir cette théorie.
Putnam, 1990b, p. 300.

La notion de stéréotype développée par Putnam opère une double césure par rapport aux théories traditionnelles de la signification lexicale et du lexique :

- traditionnellement, la signification lexicale se définit en termes de « conjonction de propriétés »⁹⁷. Putnam montre, avec les exemples des citrons verts et des tigres à trois pattes, que cette conjonction de propriétés n'est le plus souvent que la description d'un « membre normal » d'une catégorie et que cette description est hétérogène (composants de nature linguistique et extra-linguistique) ;
- définir la signification lexicale en termes de conjonction de propriétés, c'est également, relève Putnam, adopter une conception unifiée et homogène du lexique : si une description analytique en termes de propriétés rend bien compte de mots comme *célibataire* (« homme qui ne s'est jamais marié »), elle n'est pas adaptée à tous les types d'unités lexicales⁹⁸. Ce que les linguistes ont pu déduire de la thèse de Putnam, c'est que le lexique, loin d'être une « liste amorphe d'items », est hétérogène⁹⁹.

C'est parce que le lexique est doublement hétérogène, hétérogénéité d'unités (relevant d'une définition soit analytique soit stéréotypique) et hétérogénéité des composantes de la signification de ces unités (linguistiques et extra-linguistiques), que se justifie une théorie du lexique à même de distinguer des phénomènes de nature différente mais pouvant engendrer le même effet.

Une étude du lexique, dans cette approche, porte donc non plus sur les mots pris un à un mais sur des ensembles de mots regroupés en types.

Il importe donc, lorsque l'on s'intéresse à la signification lexicale comme ce qui constitue l'autonomie d'une unité lexicale, de considérer que :

- toutes les unités lexicales ne construisent pas la signification de la même façon, et qu'en ce sens il faut distinguer différents types de

⁹⁷. Putnam 1990b, p. 292-296. Putnam critique le modèle de la définition analytique, qui peut prendre, de façon réductrice, la forme suivante : « x est un citron si x est jaune, avec un goût acidulé, une peau épaisse ».

⁹⁸. « On a soutenu que la théorie qui décrit correctement le comportement de peut-être trois cents mots décrivait correctement le comportement des termes généraux qui sont des dizaines de milliers ». Putnam 1990b, p. 294.

⁹⁹. Marandin 1990, p. 289.

mots¹⁰⁰ ;

- la signification d'une unité lexicale ne correspond pas à « quelque chose d'unique et de bien circonscrit »¹⁰¹ : la notion de stéréotype, l'une des dimensions de la signification lexicale, illustre cet aspect.

Notre recherche devra donc considérer que la définition du descripteur comme « substantif »¹⁰² est trop large : il nous faudra essayer de déterminer de quels types de substantif relève le descripteur¹⁰³.

C- Exemple d'analyse de la signification lexicale vue sous l'angle du stéréotype

Pour montrer, sur un cas précis d'analyse de mots, comment le stéréotype peut être pris en compte dans la détermination de la signification lexicale, nous présentons une analyse des mots *électrophone* et *tourne-disque* issue du cadre d'étude de la signification lexicale que propose D. Corbin pour ce qu'elle nomme les « mots construits »¹⁰⁴.

Elle montre¹⁰⁵ que le sens des mots construits est étroitement associé à leur structure morphologique, « autrement dit qu'il existe un sens prédictible à partir de la façon dont le mot est construit »¹⁰⁶, ce qui n'exclut pas que des sens lexicaux différents puissent permettre de construire le même référent. Ainsi, si les deux mots *électrophone* et *tourne-disque* peuvent être utilisés pour renvoyer à la même catégorie d'objets, ils n'en suivent pas moins un parcours référentiel différent, lié aux propriétés différentes que chacun d'eux focalise : « *électrophone* décrit le mode de production du son, *tourne-disque* le fonctionnement de l'appareil »¹⁰⁷. Cette focalisation différente porte sur des traits qui « reflètent **notre connaissance stéréotypique** des objets que ce mot dénomme. [...] En conséquence il faut admettre que la langue peut construire un sens en ne sélectionnant, dans le sens d'un mot, que des **traits stéréotypiques** »¹⁰⁸.

Cet exemple issu d'un cadre d'analyse sémantique qui exploite le phénomène de stéréotypie montre comment la signification lexicale peut être ce qui permet d'assurer l'autonomie d'une unité lexicale. En effet, la signification d'un mot n'y

100. Putnam 1990b : « il y a différentes sortes de substantifs ».

101. *Id.*

102. Voir le discours normatif, présenté au § II.2 du chapitre I.

103. Voir le chapitre V.

104. « Ces mots sont construits par des opérations linguistiques, leur sens peut être calculé de façon proprement linguistique, indépendamment des catégories [référentielles] que les mots dénomment », Corbin et Temple 1994, p. 6.

105. Ses études extrêmement précises ne peuvent être reprises ici. Le cadre général de l'analyse est présenté et argumenté dans Corbin 1987.

106. Corbin et Temple 1994, p. 9.

107. *Ibid.*, p. 10.

108. *Ibid.*, p. 21. C'est nous qui soulignons.

est définie que par rapport à la langue elle-même : aucun élément extérieur - ni le concept, ni le référent - ne sont nécessaires à la détermination de la signification lexicale.

L'on voit en outre que ce qui détermine le rapprochement des termes, *électrophone* et *tourne-disque* par exemple : ce n'est pas leur signification mais leur référent. Se dessine ici un élément déterminant pour capter l'un des faits d'indexation : rappelons en effet que la synonymie documentaire se définit, dans le discours classique, contre la synonymie linguistique, en mettant en avant la ressemblance entre objets désignés par les descripteurs. Or la synonymie documentaire semble au contraire exploiter la notion de synonymie linguistique elle-même. Cependant cette exploitation ne peut apparaître que dans le cadre d'un modèle de fonctionnement de la langue qui distingue la signification lexicale (autonome, dont l'une des dimensions est stéréotypique) et la référence. Cette distinction n'étant pas opérée en indexation, les praticiens se trouvent conduits à « bricoler » de nouveaux concepts comme celui de synonymie documentaire.

On conclut sur ce point en mettant dans la perspective de l'indexation les propositions linguistiques ici présentées.

L'approche linguistique de la signification lexicale montre, nous semble-t-il, que l'indexation exploite, implicitement, par le biais du descripteur, une propriété des unités lexicales hors emploi : celle de leur autonomie, reposant en partie sur une représentation stéréotypique, qui présente des caractéristiques d'approximation. Nous dirons qu'en cela la notion de signification lexicale constitue l'un des fondements théoriques de l'indexation, dans le sens où elle fonde la possibilité de l'indexation : proposer à l'interprétation comme à l'utilisation des unités lexicales hors emploi. Une telle proposition peut paraître triviale et ne pas nécessiter une formulation en termes de fondements théoriques ; cependant nous avons vu que les approches normatives ne permettaient pas de poser la question de l'autonomie lexicale (c'est une évidence) et, encore moins, d'en proposer un traitement explicite dans le cadre documentaire.

En outre, il apparaît tout à fait nécessaire, pour comprendre la notion de synonymie en indexation, de faire apparaître une différence entre sens et référence. Nous précisons cet aspect ci-après.

II.2.2- La possibilité de désignations multiples : la question de la synonymie référentielle

D'un point de vue linguistique, la question de la synonymie ne peut être posée que dans un cadre qui distingue les niveaux, notamment celui de la signification et celui de la référence, c'est-à-dire le niveau de la langue et celui du discours, ou encore celui du lexique et de la terminologie¹⁰⁹ :

Le lexique considère les mots, la terminologie considère les choses. Il n'existe pas d'équivalence d'un mot du lexique d'une langue à un mot du lexique d'une autre langue. Mais, si l'on se place dans la perspective de la terminologie, la même classe d'objets d'un univers donné peut avoir une étiquette dans une langue et une étiquette dans une autre langue ; dès lors, la traduction devient possible, fondée sur une synonymie référentielle. Si deux termes ont la même extension dans un univers donné, on peut les considérer comme équivalents, et les traduire l'un par l'autre.
Le Guern, 1989, p. 342.

Le modèle du lexique en indexation, modèle d'utilisation et modèle de fonctionnement, en posant une homogénéité des formes lexicales, ne peut distinguer ces niveaux ; il est donc amené à créer des concepts *ad hoc* comme celui de synonymie documentaire¹¹⁰ ou de terme préférentiel¹¹¹ qui assimilent les deux types de fonctionnement du descripteur :

- son autonomie, qui correspond au niveau du « contenu » en indexation, au niveau de la signification en linguistique, au niveau du lexique dans le cadre logico-sémantique ;
- sa « dépendance », qui correspond, en indexation, à la relation qu'il établit avec les documents, à la référence en linguistique, au niveau de la terminologie dans le cadre logico-sémantique.

Dans le discours classique, le descripteur est décrit comme étant, en lui-même, de par sa « nature », un synonyme, un nom de classe d'équivalence. La signification lexicale, qui n'est pas explicitement posée, y est donc confondue avec la référence, selon un mécanisme qu'exprime ainsi Marandin sur l'exemple du couple *tête / caboche* :

¹⁰⁹. Dans le cadre du modèle logico-sémantique que propose Le Guern [1989, p. 340] : « On peut dire que le lexique concerne les mots indépendamment des choses, alors que dans la terminologie, les mots sont liés aux choses. Mais d'un côté et de l'autre, ce ne sont pas les mêmes " mots ". Ils ont bien l'air d'être les mêmes, et beaucoup de gens s'y trompent, mais l'objet " mot " pertinent pour le lexique est une réalité totalement distincte de l'objet " mot " qui appartient à la terminologie ».

¹¹⁰. Cf. Maniez 1987, Van Slype 1987, Chaumier 1988.

¹¹¹. Cf. la définition normative du descripteur [norme AFNOR Z 47-100 (1981)] : « Mot ou groupe de mots retenus dans un thésaurus et choisis parmi un ensemble de termes équivalents pour représenter sans ambiguïté une notion contenue dans un document ou une demande de recherche documentaire ». C'est nous qui soulignons.

Au regard d'une conception qui assimile signification lexicale et description de l'extension [ou référence], *tête* et *caboche* sont synonymes ; *tête* et *caboche* ne sont pas synonymes si on admet que la description de l'extension n'épuise pas la signification lexicale.
Marandin, 1992a, p. 40.

Pour capter le fonctionnement spécifique du descripteur en tant que relais textuel, c'est-à-dire comme forme lexicale susceptible de convenir à plusieurs documents, il importe donc de distinguer les niveaux de façon à pouvoir préciser celui auquel se greffe la synonymie : niveau de la langue ou niveau du discours.

On peut montrer que la notion de « synonymie » n'est pas statiquement attachée à une unité lexicale et qu'en cela la synonymie référentielle relève du discours ; sur ce point, nous reprenons l'analyse que propose Franckel à partir des deux verbes *manger* et *bouffer* :

- dans les contextes de type *On a bien mangé / On a bien bouffé*, on peut considérer qu'il y a synonymie ;
- la substitution « manger/bouffer » est moins nette avec l'exemple : *Ça se laisse manger* (paraphrasé par *C'est mangeable*) ;
- dans le cas *Il se laisse bouffer par son travail*, c'est moins « manger » qui apparaît comme un bon candidat-synonyme que les verbes « déborder » ou « accaparer ».

Cet exemple montre que si, dans certains contextes, deux termes paraissent substituables, ils n'en sont pas pour autant intrinsèquement « équivalents ».

Il y a là, comme le relève Marandin, une « illusion » dans le sens où le rapprochement de deux unités n'est « l'indice d'aucune ressemblance sémantique »¹¹².

On remarquera que le discours normatif notifie cet aspect dans la définition qu'il donne de la synonymie documentaire¹¹³, mais qu'il en fait une caractéristique distinctive du descripteur par rapport au mot de la langue. Or, comme nous le montrent les propositions linguistiques, il s'agit moins d'une spécificité du descripteur que d'une propriété de langue, et plus précisément, une propriété liée à la signification lexicale.

En effet, « multidimensionnelle » ou encore hétérogène, la signification lexicale peut être à ce titre à l'origine de manifestations diverses, pour lesquelles Marandin propose la formulation suivante :

112. Marandin 1992a, p. 46.

113. Cf. Chaumier 1978, p. 33 : « La notion de synonymie est utilisée de façon extensive dans les thésaurus sous la forme de la synonymie documentaire afin de regrouper sous un seul descripteur plusieurs termes considérés comme voisins, **bien que de signification sémantique différente** », c'est nous qui soulignons.

*Il y a de la synonymie (dans le lexique virtuel), il y a des effets de synonymie contextuelle (dans des énoncés dans certain contexte) et il n'y a pas de synonymes (dans les énoncés actuels).
Marandin, 1992a, p. 53.*

En effet, comme Le Guern, Marandin pose qu'il ne s'agit pas des « mêmes » mots selon les dimensions considérées. Marandin distingue en effet plusieurs types d'individus linguistiques, notamment les occurrences des unités lexicales (qui, dans certaines contextes, peuvent avoir des effets synonymiques) et les unités lexicales « hors emploi » (qui, à un certain niveau d'abstraction, peuvent ne pas être distinguées¹¹⁴).

Cette distinction, entre type (unité hors emploi) et occurrence (unité en contexte), est au cœur de l'explication de la synonymie documentaire que propose Le Guern¹¹⁵, à la différence près que les présupposés de la démonstration relèvent ici autant de la logique que de la linguistique. Dans son cadre, l'ensemble des unités hors emploi constitue le « lexique », dont les éléments n'ont qu'une intension* (pas d'extension, de référence) : il n'y a pas, à ce niveau, de synonymie, de traduction possible. En revanche, les unités en emploi (en discours) relèvent d'un niveau, celui de la « terminologie », où les traits référentiels permettent de concevoir l'identité référentielle et par suite la synonymie référentielle (ou documentaire¹¹⁶) ; c'est là où Marandin parle d'« effets synonymiques ». Pour Le Guern, le descripteur, en tant qu'il réfère d'une part et qu'il condense un ensemble de synonymes référentiels d'autre part, est donc un « terme », une unité de discours¹¹⁷.

Or, tout le pari de l'indexation consiste à faire fonctionner cette unité de discours qu'est le descripteur sous la forme d'une unité de langue (autonome), tentant ainsi de conjointre deux propriétés de langue différentes : en effet,

- en tant qu'élément d'une liste (liste d'un langage documentaire ou liste d'une formule d'indexation), le descripteur fonctionne comme une unité lexicale hors emploi : à ce titre, on peut dire que le descripteur est doté d'une autonomie lexicale qui vient de sa signification. Nous sommes là au niveau de la langue ;
- en tant que forme lexicale lue, interprétée et utilisée dans le contexte des documents auxquels elle est affectée, le descripteur fonctionne comme un synonyme référentiel : à ce titre, il peut permettre de

¹¹⁴. C'est le cas particulier de certains déverbaux : triche/tricherie. Marandin 1992a, p. 53.

¹¹⁵. Le Guern 1984 et 1989.

¹¹⁶. « Deux descripteurs sont synonymes s'ils ont la même référence ; il ne s'agit donc pas, dans une perspective documentaire, de synonymie lexicale, mais de synonymie référentielle », Le Guern 1984, p. 167.

¹¹⁷. Nous reviendrons largement sur ce point dans le chapitre V.

rapprocher des « objets », des documents différents. Nous sommes là au niveau du discours, de l'emploi des formes lexicales. Alors que l'approche linguistique, parce qu'elle distingue les niveaux, distingue aussi les unités, unité de langue et unité de discours, l'approche documentaire, elle, tendant à assimiler les niveaux, assimile aussi des propriétés différentes sous une même forme linguistique. En cela, si les concepts linguistiques de signification lexicale et de synonymie référentielle peuvent être vus comme des fondements théoriques de l'indexation, dans le sens où ils fondent, du point de vue d'une théorie linguistique, la pratique de l'indexation telle qu'elle s'exerce, ils ne peuvent l'être que dans le cadre d'un modèle d'utilisation du lexique : dans le modèle théorique de la langue, ces deux propriétés restent incompatibles (une forme lexicale n'est pas la « même » selon qu'elle est vue hors ou en emploi). C'est sous cet angle que nous proposons, dans le paragraphe III, un modèle d'utilisation du lexique en indexation qui, tout en étant fondé sur des propriétés linguistiques, laisse de quoi penser la torsion que réalise l'indexation.

II.3- Conclusion et résultats intermédiaires

Une lecture du discours normatif sur l'indexation permet de dégager cinq caractéristiques du descripteur et parallèlement des représentations de la langue propres au modèle objectiviste (I) :

- caractéristique 1 : le descripteur comme expression linguistique d'un concept pré-existant, stable, unitaire (« simple ») ;
- caractéristique 2 : le descripteur comme expression linguistique stable ;
- caractéristique 3 : le descripteur comme condensateur textuel ;
- caractéristique 4 : le descripteur comme expression linguistique autonome ;
- caractéristique 5 : le descripteur comme relais textuel.

Le déplacement du modèle de fonctionnement de la langue, d'un modèle objectiviste à un modèle linguistique, montre que (II) :

- les caractéristiques 1, 2 et 3 du descripteur ne peuvent être appréhendées dans un cadre linguistique. En revanche, la fonction 1 de l'indexation à la base de la formulation de ces caractéristiques peut, elle, être analysée dans le cadre d'un modèle linguistique en termes d'effet d'interprétation : effet obtenu sur la base des

descripteurs qui fonctionneraient comme thèmes de discours.

Cette hypothèse du descripteur comme thème de discours conduit à chercher comment l'indexation peut permettre à un utilisateur de construire ces thèmes de discours : nous faisons l'hypothèse que l'indexation met en œuvre une stratégie de regroupement et d'exposition des documents qui permette de telles constructions (*infra*, chapitre IV) ;

- les caractéristiques 4 et 5 du descripteur, qui découlent de la seconde fonction assignée à l'indexation, rencontrent, elles, des propriétés de langue : le descripteur peut être utilisé comme un accès autonome parce que, en tant qu'unité lexicale hors emploi, il est pourvu d'une signification lexicale ; le descripteur peut être utilisé comme synonyme parce que, en tant qu'unité de discours, il peut renvoyer à des objets référentiellement différents.

Ces aspects de la signification lexicale et de la synonymie référentielle en jeu dans l'indexation *via* son résultat (le descripteur), pourraient constituer directement des fondements théoriques s'ils étaient utilisés, dans l'indexation, d'une façon qui les maintienne distincts. Or, l'indexation attribue à une même forme lexicale des propriétés distinctes, « incompatibles ». C'est ainsi qu'il devient nécessaire d'essayer de concevoir un modèle d'utilisation du lexique qui pense ces différents niveaux et les articule.

III- Reformulation du modèle d'utilisation du lexique en indexation

Nous avons montré en I que le modèle d'utilisation du lexique en indexation reposait sur des représentations de la langue non valides du point de vue de la théorie linguistique et non complètes d'un point de vue descriptif. Nous avons vu en II qu'un modèle linguistique pouvait proposer des explications du fonctionnement des mots en indexation. Si nous adoptons les représentations linguistiques de la langue que nous avons présentées en II, que reste-t-il de l'indexation et du descripteur tels qu'ils ont été abordés en I par le biais des normes ? À quelles définitions de l'indexation (III.1) et du descripteur (III.2) arrivons-nous ?

III.1- L'indexation dans le cadre d'une approche linguistique du lexique

Dans le nouveau cadre de fonctionnement du lexique adopté, la définition normative de l'indexation n'est plus valide.

On se souvient que l'indexation est définie dans la norme comme un processus réalisé en deux phases : une phase d'analyse conceptuelle (analyse de contenu) et une phase de représentation linguistique (traduction). Ce processus en deux phases semble devoir être mis en cause (III.1.1) ; par suite, cette remise en cause conduit à mettre à distance le rôle des mots en indexation (III.1.2).

III.1.1- Mise en cause du processus en deux phases

Il apparaît que la description de l'indexation en deux phases est artificielle : la reconnaissance de concepts d'une part et leur codage par des descripteurs d'autre part ne valent que dans le cadre d'un modèle objectiviste de la langue. Hors de ce modèle, le processus en deux phases ne peut être vu que comme une décomposition méthodologique établie *a posteriori* à des fins didactiques, ou du moins supposées didactiques.

La faillite de la description d'un processus en deux phases, si elle ressort d'une approche théorique au terme d'une confrontation entre modèles de langue, a pu être mise en évidence sur la base d'expérimentations menées dans le cadre des sciences cognitives. Ainsi Sylvie Bruxelles [1991], qui a étudié, d'un point de vue linguistique et psychologique, les codages réalisés par plus de 150 sujets à partir de deux nomenclatures¹¹⁸, remarque-t-elle que :

¹¹⁸. Bruxelles 1991, p. 171-186 : l'expérimentation porte sur des documents juridiques (40

- (i) le codage, la classification, ne relève pas de la traduction (mot à mot, concept à concept, ou encore concept à mot) mais de l'intertextualité, de la confrontation de deux textes. En effet, en montrant les contraintes linguistiques et cognitives exercées par les nomenclatures sur le choix des postes de classement, Bruxelles note que l'instrument d'indexation constitue le texte d'arrivée qui permet de lire le texte-source (le document à indexer). Outre que le codeur, contrairement au traducteur, dispose d'emblée des deux textes (source et cible), l'orientation de leur activité est inverse : la traduction opère de la source à la cible alors que l'indexation lit la source en fonction de la cible ;
- (ii) en ce sens, le codage ne suppose pas une analyse de contenu préalable, entendue comme une « décomposition analytique des propriétés des objets manipulés. Les associations sémantiques se forment plutôt par agrégation à partir de foyers-repères où se nouent des “reconnaisances” »¹¹⁹. En ce sens, les nomenclatures apparaissent comme des « principes de construction de l'information »¹²⁰ et non comme de simples outils de transmission d'une information pré-existante. C'est pourquoi le document, le texte classé, apparaît comme un construit, ou un reconstruit¹²¹.

Ces résultats d'analyse d'activités de codage, succinctement rapportés ici, confirment certaines de nos conclusions, établies sur un plan plus formel :

- le processus de l'indexation peut être dégagé des notions d'analyse de contenu et de traduction (point (ii)) ;
- sur le plan de l'utilisation des mots (de l'attribution des descripteurs), l'indexation semble se réaliser en une seule phase (point (i)), la fonction de fourniture d'accès supplantant celle de représentation du contenu ;
- l'indexation apparaît comme une opération de construction du document (point (ii)).

L'utilisation des mots en indexation ne semble donc pas relever de la traduction, ni de l'analyse de contenu, mais plutôt de la construction de l'information qui se ferait *via* l'établissement d'accès aux documents et *via* la construction des documents

assignations et requêtes issues de juridictions) indexés à l'aide de deux nomenclatures (plus précisément à l'aide de deux versions de la *Nomenclature des affaires civiles*, celle de 1981 et celle de 1988) par trois groupes d'individus : 92 greffiers, 42 élèves de l'École nationale de la Magistrature, 30 étudiants non juristes.

¹¹⁹. *Ibid.*, p. 182.

¹²⁰. *Id.*

¹²¹. *Ibid.*, p. 183 : Bruxelles note que les codeurs « mettent en œuvre des procédures de reconstruction d'objets, dominées par des effets de contexte ».

eux-mêmes. D'autres mécanismes langagiers sont alors à postuler. Pour ce qui est de l'indexation « contrôlée », Bruxelles suggère que les relations d'intertextualité sont à prendre en compte dans le processus de l'indexation. Nous ferons l'hypothèse plus large que le principe de la confrontation textuelle est à l'œuvre dans tout type d'indexation (chapitre IV).

III.1.2- Mise à distance des mots en indexation

La mise en cause de l'indexation comme traduction d'un contenu pré-existant induit des conséquences à différents niveaux, qui conduisent à mettre à distance le rôle des mots en indexation :

- *Conséquences sur la fonction de l'indexeur*

L'indexeur ne peut plus être vu comme un traducteur, pris dans un rapport paradoxal avec le langage, considéré sous la double facette du « coupable » et du « rédempteur ». On retrouve ici la fin du mythe du « médiateur », que l'analyse d'autres activités de diffusion de connaissances, comme la vulgarisation scientifique, a pu mettre en valeur¹²² ; ainsi celle menée par Jacobi :

La science ne serait pas comprise avant tout parce qu'elle se parle dans une langue ésotérique. [...] Cette idéologie impose la figure du médiateur qui, par sa compétence de traduction, parvient à rétablir la communication. [...] Dans cette perspective, le langage apparaît sous le double visage du coupable et du rédempteur.
Jacobi, 1987, p. 26.

Sur ce point, il apparaît que le travail de l'indexeur doit porter moins sur une manipulation de mots (trouver les « bons » mots qui permettront aux utilisateurs de communiquer avec des auteurs) que sur une mise en contexte des mots susceptible de permettre cette communication.

- *Conséquences sur l'approche de la langue*

Le lexique, l'utilisation de mots en indexation, n'engage pas forcément, comme l'indique le modèle linguistique, une approche strictement lexicaliste de la langue ; il peut, au contraire, ouvrir la voie au discours, à l'analyse des discours. C'est une tout autre appréhension du lexique qui peut alors être adoptée :

Aussi faut-il voir dans le lexique moins une donnée contraignante, dont l'emploi serait soumis au seul principe d'adéquation référentielle, qu'un ensemble de dispositifs extrêmement malléables, continuellement travaillés et retravaillés dans et par le discours.
Apothéloz et Reichler-Béguclin, 1995, p. 241.

¹²². On revient sur ce point dans le chapitre IV, § I.3.

Si elle s'exprime prioritairement par les mots, l'indexation n'est donc pas uniquement concernée par eux. Nous aurons donc à dégager la dimension discursive de l'indexation.

- *Conséquences sur l'approche du langage documentaire*

La notion de langage documentaire est amenée, elle aussi, à être repensée. Elle se donne traditionnellement pour un métalangage, substituant à des noms d'autres noms (traduction mot₁ -> mot₂ dans le schéma classique de l'indexation). Or, dans le cadre du modèle de fonctionnement de la langue adopté, « le métalangage est une illusion logique : il n'y a pas de noms de noms, mais seulement des noms de gestes, d'événements, de choses »¹²³ ; la notion de métalangage ne tient que si l'on ne distingue pas les niveaux, ici celui des mots et celui des choses¹²⁴. L'enjeu du langage documentaire n'est donc plus celui d'être un outil de représentation du contenu mais plutôt celui d'être un outil de construction du contenu (Bruxelles 1991, *supra*).

Il apparaît que le déplacement du modèle de fonctionnement du lexique tel qu'on le propose opère un renversement des données : ce n'est plus, en indexation, la représentation du contenu d'un document qui permet l'accès à l'information d'un fonds documentaire, c'est plutôt la fourniture d'accès aux documents qui permet de construire une représentation du contenu d'un document.

L'enjeu de l'indexation se pose alors moins dans le cadre du lexique (niveau de la signification lexicale, du contenu) que dans celui de la référence (niveau des référents et des discours). L'indexation rejoint par là des problématiques plus générales, communes à d'autres pratiques :

Le problème n'est donc plus de se demander comment l'information¹²⁵ est transmise ou comment des états du monde sont représentés de façon adéquate, mais de se demander comment les activités humaines, cognitives et linguistiques, structurent et donnent un sens au monde.

Dubois et Mondada, 1995, p. 276.

Dans ce cadre, l'indexation n'a plus à faire se rencontrer des « mots » mais plutôt des « mondes », ceux des auteurs et ceux des utilisateurs, et ceci par le biais de mots, les descripteurs. L'approche du descripteur en indexation est donc nécessairement au moins double : elle se mène à la fois du côté du lexique et du côté de la référence.

¹²³. Berrendonner 1981, p. 132.

¹²⁴. *Id.*, note 19 : « Je me refuse d'admettre que les objets situés à deux niveaux contigus soient des objets du même ordre : pour moi, la référence relie toujours une chose et un nom, et la chose ne saurait être un nom ».

¹²⁵. Le terme « information » est à prendre ici dans l'acception que lui donnent les sciences cognitives.

III.2- Le descripteur dans le cadre d'une approche linguistique du lexique

Notre approche de l'indexation nous conduit à considérer le descripteur comme un « accès documentaire », c'est-à-dire comme une forme lexicale susceptible de conjointre des propriétés linguistiquement incompatibles :

- pour constituer un accès autonome, le descripteur doit être appréhendé comme une unité lexicale hors emploi, pourvue d'une autonomie : sa signification lexicale, qui révèle une certaine stabilité (stabilité sémantique de nature stéréotypique, largement sous-déterminée) ;
- pour constituer un accès « multiple » (à plusieurs documents), le descripteur doit être une unité de discours pour fonctionner, aux yeux des utilisateurs, comme un thème de discours susceptible de construire des effets de synonymie référentielle.

Sur un plan linguistique, les deux niveaux de langue et de discours sont tenus pour distincts et l'individu de langue n'est pas le même que l'individu de discours. Or, l'indexation a besoin d'utiliser les mots sur les deux plans en même temps : dans le cadre de quel modèle d'utilisation du lexique en indexation peut-on rendre compte de cette dualité du descripteur, à la fois type ET occurrence ?

On se propose, à la suite de Michel Le Guern et au travers de sa propre analyse, d'appréhender la relation type/occurrence dans le cadre proposé par Peirce : Le Guern propose en effet une lecture de Peirce qui permet de voir comment le descripteur peut fonctionner à un double niveau de langue et de discours¹²⁶.

Il ne s'agit pas, pour nous, d'exposer le détail de cette approche : nous nous contenterons, d'une part, d'indiquer quelques notions générales pour situer la problématique de Peirce et nous présenterons, d'autre part, certains des concepts de Peirce dégagés par Michel Le Guern comme pertinents en matière d'explicitation du mécanisme d'indexation.

¹²⁶. La démarche de Le Guern [1991a] s'inscrit dans le cadre d'une entreprise de « désintrication » des niveaux, non distingués dans le discours classique en raison de l'absence de référentiel théorique : « Il est fâcheux que l'on ait appelé les descripteurs des " mots-clés ", ce qui laisse croire que ce sont des mots, des unités lexicales. Or, dans les pratiques documentaires les plus courantes, l'indexation vise les objets, les référents, et non les signifiés. Cette distinction revêt d'autant plus d'importance qu'elle se heurte plus fortement à la conception naïve — généralement répandue chez les utilisateurs de la langue qui ne sont ni linguistes ni logiciens — des relations entre les mots et les choses. Cette conception, d'après laquelle les mots, en tant qu'unités lexicales et préalablement à toute insertion dans le discours — ou tout au moins certains d'entre eux —, désigneraient directement les choses, peut être appelée l'illusion du substantif. Cette illusion remonte au moins à Aristote, et il n'était sans doute pas possible d'y résister tant qu'on n'apercevait pas de manière nette l'opposition langue / parole », Le Guern 1991a, p. 23.

III.2.1- Brève présentation du modèle de Peirce

Rappelons, pour commencer, que la théorie sémiotique de Peirce ne s'appuie pas sur des présupposés de nature linguistique ; relevant d'une réflexion d'ordre phénoménologique, elle explore plutôt la logique des relations¹²⁷. Les travaux de Peirce sont néanmoins considérés comme étant à l'origine de la distinction, formalisée ensuite par Morris, entre les trois « niveaux » de l'analyse linguistique, syntaxe, sémantique et pragmatique¹²⁸.

Les principales notions que manipule Peirce sont le signe et la sémiosis, dont la problématique peut s'énoncer de la façon suivante :

Toute chose, tout phénomène, aussi complexe soit-il, peut être considéré comme **signe** dès qu'il entre dans un processus sémiotique, c'est-à-dire dès qu'un interprète le réfère à autre chose.

Everaert-Desmedt, 1990, p. 25.

La « **sémiosis** » ou la production de la signification est un processus triadique qui met en relation un representamen, un objet et un interprétant.

Everaert-Desmedt, 1990, p. 26

De façon schématique, on entend par :

- representamen : une chose qui représente une autre chose ;
- objet : une entité physique ou mentale ;
- interprétant : une action de médiation entre le representamen et l'objet¹²⁹.

Chacun de ces trois éléments se subdivise encore lui-même en trois catégories :

- selon l'ordre dont il relève¹³⁰, un representamen sera un qualisigne (priméité), un sinsigne (secondéité) ou un légisigne (tiercéité) ;
- selon le type de relation que l'objet entretient avec le representamen, l'objet sera un icône (relation de similarité), un indice (relation de contiguïté contextuelle) ou un symbole (relation arbitraire, provenant d'une règle, d'une loi, etc.) ;
- selon le type de règle qui renvoie le representamen à son objet, l'interprétant sera un rhème (règle reposant sur les caractères de

¹²⁷. Voir Deledalle *in* Peirce 1978, p. 212 : « La sémiotique est, selon Peirce, un autre nom de la logique : “ la doctrine *quasi* nécessaire ou formelle des signes ” ».

¹²⁸. On peut trouver une présentation générale de l'approche de Peirce dans Everaert-Desmedt 1990. Nous nous sommes en grand partie appuyée sur cette présentation dans le paragraphe qui suit.

¹²⁹. Peirce précise que « l'interprétant n'est pas l'interprète mais le moyen que celui-ci utilise pour effectuer son interprétation », *in* Everaert-Desmedt 1990, p. 40.

¹³⁰. Peirce distingue trois ordres : la priméité (le possible), la secondéité (le réel) et la tiercéité (la loi), qui constituent les principes de subdivision en catégories à la fois dans le representamen, l'objet et l'interprétant, voir Everaert-Desmedt 1990, p. 48 notamment.

l'objet seulement), un dicisigne (règle reposant sur l'existence de l'objet) ou un argument (règle exploitant l'objet en tant que signe).

Pour qu'il y ait signe, il faut, comme nous l'avons vu précédemment, que s'établisse un « processus triadique », c'est-à-dire une relation entre un representamen, un objet et un interprétant. En fonction des trois trichotomies ci-dessus présentées, on devrait avoir 27 types de signe possibles ; mais, ces trichotomies étant fondées sur des « ordres », exprimant une hiérarchie, certaines des combinaisons possibles ne sont pas valides. C'est ainsi qu'on retient généralement dix modes de fonctionnement de la signification (ou « signes »)¹³¹.

III.2.2- Approche du descripteur dans le modèle de Peirce

C'est à partir du cadre posé par Peirce que Michel Le Guern¹³² propose de faire « voir » la différence et la ressemblance entre mot de la langue et descripteur¹³³ : il situe le point de distinction sur la façon dont chacun de ces « signes » gère la relation type/occurrence, c'est-à-dire la relation representamen/objet ou encore mot/chose.

Ainsi Michel Le Guern explique-t-il que le descripteur peut s'analyser comme un « légisigne indiciaire rhématique », tandis que le mot de la langue est, lui, un « légisigne symbolique rhématique ». Les deux types de signe sont des « légisignes »¹³⁴, dont la propriété est de ne « pouvoir agir qu'en se matérialisant dans des insignes qui constituent des répliques ». En effet, le légisigne est un « type général », accessible uniquement par le biais de ses occurrences :

En soi, un signe est soit une apparence, ce que j'appelle un *qualisigne*, soit un objet ou événement individuel, ce que j'appelle un *sinsigne* (la syllabe *sin* étant la première syllabe de *semel*, *simul*, *singulier*, etc.), soit un type général, ce que j'appelle un *légisigne*. Comme nous employons le terme « mot » dans la plupart des cas, quand nous disons que « le » est un « mot », que « un » est un autre mot, un « mot » est un légisigne. Mais quand nous disons d'une page d'un livre qu'elle a

¹³¹. Ces dix signes sont d'après Everaert-Desmedt 1990, p. 94 :

qualisigne iconique rhématique	légisigne indiciaire rhématique
sinsigne iconique rhématique	légisigne indiciaire dicent
sinsigne indiciaire rhématique	légisigne symbolique rhématique
sinsigne indiciaire dicent	légisigne symbolique dicent
légisigne iconique rhématique	légisigne symbolique argumental

¹³². Le Guern 1984, 1989, 1991a.

¹³³. La démonstration menée par Le Guern a pour objectif de déplacer l'appréhension du descripteur d'une sémantique lexicale et à une sémantique discursive ; dans cette perspective, ce qui est opposé, c'est le mot du lexique (appelé, dans ce cadre, « mot de la langue ») et le mot du discours (en l'occurrence le descripteur). Si l'on veut bien s'abstraire des différences de terminologie, on conviendra que les vues exprimées ici convergent avec celles précédemment exposées dans le § II.

¹³⁴. « Le légisigne est un signe dont le fondement est une loi. Une loi est établie *a priori*, par convention, décision arbitraire ; ou *a posteriori*, par habitude ». C'est pourquoi il a une « identité bien déterminée », Everaert-Desmedt 1990, p. 51.

deux cent cinquante “ mots ” dont vingt sont des “ le ”, le “ mot ” est un sinsigne. Un sinsigne qui renferme ainsi un légisigne, je l'appelle une “ réplique ” du légisigne.

Peirce, cité *in* Deledalle, 1990, p. 85.

Mais, remarque Michel Le Guern, selon qu'il est « indiciaire » ou « symbolique », le légisigne n'établit pas la même relation entre type et occurrence¹³⁵ : en tant qu'indice, le descripteur ne fait que **désigner** son objet, alors que le mot de la langue, étant symbole, signifie son objet par l'intermédiaire d'un **interprétant**. En effet, le rôle de l'indice est d'assurer la référence et de ne faire que cela : « sa fonction est pragmatique et non sémantique »¹³⁶ ; par opposition, le symbole établit une relation indirecte avec son objet, par l'intermédiaire de l'interprétant :

Un symbole est un signe qui perdrait le caractère qui en fait un signe s'il n'y avait pas d'interprétant. Exemple : tout discours qui signifie ce qu'il signifie par le seul fait que l'on comprenne qu'il a cette signification.

Peirce cité *in* Deledalle 1990, p. 86.

En simplifiant, on pourrait dire que, dans le cas du légisigne indiciaire rhématique, la relation type/occurrence est de type référentiel ; elle est de nature « sémantique » dans le cas du légisigne symbolique rhématique.

Cependant, comme le précise Le Guern, « le mot de la langue est également l'interprétant du descripteur »¹³⁷ : en effet, l'indice a simplement pour objet de montrer un objet, « il appartient au symbole d'en parler »¹³⁸.

Autrement dit, en tant qu'indice, le descripteur est une unité de discours : en ce sens il peut, *via* ses occurrences, désigner différents objets singuliers, et permettre de constituer, de texte en texte, un thème de discours. En tant qu'indice toujours, le descripteur entretient une relation avec l'unité lexicale hors emploi (son interprétant) : la signification lexicale que se voit alors attribué le descripteur, notamment par sa sous-détermination référentielle, permet de le doter d'une certaine stabilité sémantique, qui ne préjuge en rien de son instabilité référentielle.

C'est par l'établissement de ces deux fonctions du descripteur en tant qu'indice (« montrer » et « dire ») que Michel Le Guern propose de voir le descripteur comme un nom propre¹³⁹ tel qu'il est entendu dans la typologie de Peirce¹⁴⁰. Perçu

¹³⁵ Le Guern 1984, p. 165.

¹³⁶ Everaert-Desmedt 1990, p. 64.

¹³⁷ Le Guern 1984, p. 165.

¹³⁸ Deledalle *in* Peirce 1978, p. 235.

¹³⁹ « Si le descripteur est le signe de ses occurrences dans le corpus, ce n'est pas comme s'il en était en quelque sorte la reproduction photographique. Il en est plutôt le “ nom propre ” ». Le Guern 1984, p. 166.

¹⁴⁰ Voir Thibaud 1989, p. 381 : « Une expression est un nom propre si et seulement si il est possible de l'introduire comme index d'un objet individuel, de telle sorte que ce nom peut être

comme légisigne indiciaire rhématique¹⁴¹, le nom propre constitue pour Peirce un véritable paradoxe, dans le sens où ce qu'il représente ne se réduit jamais à ce qui en est dit :

Affronté à la nécessité de désigner des événements singuliers par le moyen des termes ayant une signification générale, le langage trouve, dans le nom propre, un moyen de dépasser embrayeurs et descriptions : tandis que les premiers montrent sans rien dire (index purs) et que les secondes disent sans montrer (symboles iconiques purs), le nom propre en tant qu'articulation symbolico-iconico-indexique est ce qui relie un dire à une monstration. On peut donc dire qu'il est la voie d'accès privilégiée à l'individuel.

Thibaud, 1989, p. 386.

En suivant la proposition de Le Guern, le descripteur, appréhendé comme un nom propre au sens que Peirce attribue à ce mot, apparaît donc comme une unité qui « désigne un objet que n'épuise aucune description »¹⁴² : la signification lexicale du descripteur n'épuise pas, en effet, sa référence.

Ce modèle du descripteur comme nom propre de ses occurrences, proposé par Michel Le Guern sur la base de la théorie peircienne du signe, permet, nous semble-t-il, de comprendre comment l'indexation peut articuler signification lexicale et synonymie référentielle et conjointre ainsi la stabilité de la signification avec l'instabilité de désignation.

En outre, ce modèle dégage la notion de nom propre, que nous avons par ailleurs déjà vu apparaître, et précise l'usage que l'on peut en faire dans le cadre d'un modèle d'utilisation de la langue : si le descripteur fonctionne comme un nom propre, il s'agit là d'une caractéristique de fonctionnement et non d'une propriété de « nature ». Le descripteur n'a pas à être un nom propre sur un plan linguistique mais doit se trouver doté de ses particularités de fonctionnement, notamment celle de désigner des objets individuels. Cette propriété lui vient de sa catégorie grammaticale : le nom propre est un groupe nominal. Le descripteur devra donc pouvoir être un groupe nominal, en discours, en même temps qu'une unité lexicale, hors emploi : c'est sur cette base que nous développerons notre approche du descripteur dans le chapitre V.

Enfin, cette approche du descripteur « fonctionnant comme un nom propre » met au jour l'enjeu de l'indexation : mettre à disposition des utilisateurs des procédés de désignation stable. Ces procédés ne relèvent pas uniquement des mots eux-mêmes,

utilisé comme symbole dans des situations différentes de celle où l'objet est présent et indexiquement reconnu ».

¹⁴¹. Par opposition au nom commun, compris comme un légisigne symbolique rhématique, qui, en tant que symbole, ne peut en lui-même « identifier les choses », Deledalle *in* Peirce 1978, p. 165.

¹⁴². Thibaud 1989, p. 384.

du lexique : nous aurons à introduire la dimension du discours pour comprendre comment l'indexation peut réaliser un tel pari.

IV- Conclusion du chapitre

Pour peu apparente qu'elle soit dans le discours normatif, la dimension lexicale de l'indexation nous paraît essentielle à faire apparaître. On peut en effet, par elle, d'une part, distinguer ce qui dans l'indexation relève des effets d'utilisation de la langue et ce qui relève des propriétés de la langue elle-même et, d'autre part, approcher la finalité de l'indexation dans des termes qui ne soient pas circulaires.

Sur ces deux points, ce chapitre a permis d'obtenir les résultats suivants :

(i) Distinction des faits et des effets

• Analyse des effets et conséquences sur la poursuite de la recherche

Nous sont apparus comme pouvant être des effets d'interprétation les aspects de l'indexation relevant de l'analyse et de la représentation du contenu. Cette approche en termes d'effets nous conduit à poser l'hypothèse que le descripteur doit pouvoir fonctionner, aux yeux d'un utilisateur, comme un thème de discours. Une telle hypothèse nécessite de considérer l'indexation comme un processus particulier d'organisation de textes (notion d'interdiscours en indexation). D'où la nécessité de poursuivre la recherche en abordant l'indexation sous l'angle du discours¹⁴³. L'analyse du thème de discours ne relève pas, en effet, d'une sémantique lexicale, mais d'une sémantique discursive.

• Analyse des faits et première approche des fondements théoriques de l'indexation

En proposant à l'utilisation des « mots » isolés, l'indexation nous est apparue comme exploitant une propriété des unités lexicales hors emploi : celle de leur autonomie lexicale, que l'on a proposé d'approcher sous l'angle de la signification lexicale. Parallèlement, nous avons relevé que l'indexation exploitait aussi le fonctionnement synonymique des unités lexicales en discours : ce n'est qu'en discours qu'une unité lexicale peut désigner des « objets » différents.

Si l'indexation exploite des propriétés spécifiques aux unités lexicales de la linguistique, elle les exploite d'une façon qui les rend incompatibles au regard d'une théorie de la langue. Si les propriétés d'autonomie lexicale et de synonymie référentielle constituent des fondements théoriques de l'indexation, ce ne peut être que dans le cadre d'un modèle d'utilisation de la langue qui permet de les articuler de façon non contradictoire. C'est ainsi que nous avons proposé de commencer à bâtir ce modèle d'utilisation : le descripteur peut être considéré comme le nom propre de ses occurrences dans un corpus. Se dégage ici une première piste : le descripteur, qui doit pouvoir fonctionner en contexte comme un groupe nominal,

¹⁴³. Voir le chapitre IV.

doit être hors contexte une unité lexicale susceptible d'apparaître dans un groupe nominal. Le modèle d'utilisation de la langue en indexation, ici à ses débuts, sera plus amplement spécifié dans la seconde partie de cette recherche.

(ii) Approche de la finalité de l'indexation

L'indexation nous est apparue comme une pratique destinée moins à **transmettre** de l'information, suivant une chaîne linéaire qui irait des auteurs aux utilisateurs, qu'à permettre de **construire** cette information. L'indexation ne se donne plus comme une opération qui doit déterminer, à partir de l'analyse d'un document, l'information dont l'utilisateur pourra avoir besoin. Elle agit sur un autre plan. Rapprochée de la vulgarisation scientifique, l'indexation se conçoit comme une pratique mettant en œuvre des **stratégies d'exposition** des textes, agissant plus sur les conditions d'interprétation que sur l'interprétation elle-même. En cela, elle est directement concernée moins par les « mots » eux-mêmes que par leur « mise en contexte ».

Remarquons sur ce point que l'on peut rendre compte de la finalité de l'indexation sans la concevoir en termes de recherche documentaire proprement dite. L'horizon de l'utilisateur reste présent dans notre approche mais l'utilisateur est perçu plus sous l'angle de son « rôle » (rôle d'interprète) que sous l'angle de sa « nature » (un individu singulier en quête d'information particulière).

À plusieurs reprises, dans ce chapitre, a été évoqué l'« autour » des mots ou plus précisément ce à quoi ils réfèrent : des discours, des objets, des « choses », des contextes, etc. La question du lexique en indexation appelle en effet nécessairement celle de la référence. Nous nous proposons de préciser cet aspect de la référence en indexation dans le chapitre suivant.