

Conclusion

L'indexation documentaire est le plus souvent appréhendée par les professionnels de l'information et de la communication sous la seule dimension instrumentale. C'est en se focalisant sur la question « à quoi sert-elle ? » (en l'occurrence à la recherche d'information) que l'indexation se définit, se laisse décrire et évaluer.

Le choix d'un tel angle d'approche se comprend aisément dans le cadre des pratiques professionnelles où dominant les exigences de performance. Toutefois, il se révèle réducteur dès lors que l'indexation devient l'objet d'autres problématiques, notamment celles de l'évolution des technologies de l'information. Le point de vue « instrumental » ne permet pas alors de discuter l'indexation dans ses spécificités. Il conduit de proche en proche à la faire apparaître comme une simple « technique », parmi d'autres techniques possibles : l'objet « indexation » proprement dit tend à disparaître des problématiques documentaires ; du moins, il n'est plus considéré comme central.

Mettant à distance la finalité ainsi entendue de l'indexation, cette recherche s'est intéressée à dégager les spécificités, les caractéristiques, les propriétés de l'indexation. Pour ce faire, elle a cherché à établir les fondements théoriques de l'indexation. Ceux-ci ont été appréhendés partiellement, du point de vue d'une théorie linguistique.

I- Le point de vue linguistique sur l'indexation : une tentative de distinction

Le point de vue linguistique sur l'indexation permet de dégager les représentations sous-jacentes de la langue et du langage à l'œuvre dans les pratiques professionnelles. Apparaissent alors des confusions entre faits de langue et faits d'indexation. Des principes propres à l'interprétation des discours comme à celle des unités lexicales se donnent comme des objectifs singuliers de l'indexation.

Ainsi la notion de « représentation du contenu d'un texte par des mots » devient-elle définitoire du processus de l'indexation, engageant la construction d'instruments spécifiques comme les langages documentaires, alors que le point de

vue linguistique montre qu'il s'agit là d'un principe de construction thématique qui, se dégageant au niveau des discours, dépend de la langue elle-même. De même, la notion de « stabilité », posée par les praticiens en termes de « continuité conceptuelle » et envisagée sur le seul plan lexical, peut être réinterprétée par l'examen des propriétés de langue en jeu dans la construction de la référence. De ce point de vue, la stabilité référentielle se comprend en termes d'effet d'interprétation : l'indexation apparaît alors comme une pratique destinée à construire de tels effets de stabilité référentielle.

Dès lors que le « monde » n'est plus pensé comme une donnée stable, l'indexation n'a plus pour fonction de « transmettre » une information qui lui préexisterait ; elle doit plutôt disposer d'éléments de nature à permettre la construction de l'information.

Par la confrontation systématique entre modes de représentation de la langue (des linguistes d'une part et des praticiens d'autre part), on parvient à distinguer les deux niveaux - du fonctionnement de la langue et de l'utilisation de la langue - en jeu dans l'indexation. Dès lors, on est en mesure de pouvoir spécifier ce qui caractérise en propre l'indexation. On peut rapporter aux propriétés de la langue elles-mêmes une partie de la pratique d'indexation ; on peut faire émerger parallèlement les zones où crucialement la pratique d'indexation se constitue comme telle.

Sur ces deux points, on peut montrer que :

- la viabilité de l'indexation repose sur l'exploitation des propriétés linguistiques suivantes : la signification lexicale, la synonymie référentielle, la construction de la référence discursive, la rigidité désignative. Cet ensemble de propriétés linguistiques fonde la possibilité de l'indexation, mais il ne garantit ni sa réalisation effective ni sa performance ;
- sur la base de ces propriétés de langue, l'indexation peut se donner un projet spécifique qui, du seul point de vue linguistique, se laisse appréhender par la notion de construction de thèmes discursifs. La spécificité de l'indexation peut se définir en fonction d'un tel projet. Elle se fonde alors sur les caractéristiques de la construction des thèmes de discours : les thèmes de discours se construisent à travers plusieurs discours (notion d'interdiscours), et, plus précisément, à travers la saisie de plusieurs objets de discours (modèle des chaînes de référence). Le thème de discours relève en cela d'une lecture, d'une interprétation : il ne préexiste pas aux textes, dont il ne peut, en conséquence, être « extrait ». Le thème

de discours se matérialise en outre sous une forme linguistique spécifique : le groupe nominal.

II- Spécificité de l'indexation : un espace de discours

Distingués des faits de langue, les faits d'indexation font apparaître leur singularité à deux niveaux (le document et le descripteur) qui relèvent tous deux d'une dimension discursive.

II.1- Le discours documentaire

L'indexation apparaît d'abord comme un processus de création d'espaces documentaires spécifiques. Les questions qu'elle doit résoudre se dédoublent alors en deux problématiques :

- comment construire une collection documentaire qui permette une multiplicité de thématisations ?
- comment rendre accessible la collection documentaire ainsi constituée ?

Pour que ces deux questions puissent recevoir un cadre de réponse général dans lequel la particularité des pratiques puisse s'exprimer, on peut essayer de préciser le niveau de leur intervention, c'est-à-dire dégager un système de contraintes à partir duquel pourront être établies des règles spécifiques.

Concernant la constitution d'une collection documentaire, le principe directeur peut être défini à partir des conditions de l'existence et de la co-existence des textes eux-mêmes : à l'horizon théorique de l'indexation se place alors la notion d'« interdiscursivité ». L'examen des conditions qui déterminent l'existence des discours peut fournir des éléments de nature à guider la construction de l'interdiscours en indexation.

La problématique de l'accès à un espace spécifique de documents rejoint celle de la stabilité référentielle et peut être pensée à travers la réflexion menée en logique sur l'identité d'un objet. Cet arrière-plan théorique pose la question de la « reconnaissance » d'un « même » objet dans des termes sensiblement différents de ceux de l'appariement, de la rencontre entre expressions linguistiques semblables. La problématique de l'accès à une collection documentaire ne se réduit plus au choix du mot juste. L'approche logique montre en effet que l'identité d'un

objet ne relève pas d'une mesure de ressemblance : elle suppose plutôt une construction qui permette à un objet de recevoir, tout en restant lui-même, de nouvelles propriétés. L'indexation apparaît alors comme cette construction susceptible de contraindre le parcours interprétatif des utilisateurs à travers l'espace documentaire qu'elle a constitué.

En tant qu'espace d'organisation spécifique des documents, l'indexation se laisse décrire comme un processus mettant en jeu deux types de stratégie : une stratégie d'exploration des sources et une stratégie d'exposition des documents. Ces deux stratégies s'établissent dans un objectif bien déterminé : la construction de thèmes de discours. Cet objectif relève moins de la recherche d'information proprement dite (le « contenu » des thèmes de discours construits par les utilisateurs n'intéresse qu'indirectement l'indexation) que d'un principe d'organisation des discours (sur quelles bases rapprocher des documents ? Les rendre complémentaires ? Les rendre accessibles ?).

On peut donc définir l'indexation non plus en fonction de son objectif (à quoi sert-elle ?) mais en fonction de ses objets : qu'est-ce que des textes, des documents ? Quels liens existent entre eux ? Comment permettre l'exploitation de ces liens ?

En prenant en compte la spécificité des objets qu'elle manipule, l'indexation peut trouver, dans ses propres objets, un « mode d'emploi » qui lui garantisse une certaine régularité de procédure. C'est la prise en compte de la condition d'existence des textes dans la stratégie d'exploration des sources ; c'est l'exploitation de la condition d'existence des objets de discours dans la stratégie d'exposition des documents.

II.2- Le descripteur dans le discours documentaire

Dans le cadre de l'indexation comprise comme organisation spécifique de documents, la problématique du descripteur demande une nouvelle formulation. Le descripteur n'apparaît plus comme ce qui rapproche *a posteriori* des documents différents ; il se donne plutôt comme un élément qui permet de « circuler » dans un espace documentaire conçu *a priori* comme homogène. Cet élément doit présenter la caractéristique de pouvoir participer à la construction de thèmes discursifs. Cette contrainte détermine son rôle comme sa morphologie.

Le rôle du descripteur s'exprime clairement d'un point de vue logique : le descripteur doit établir une double relation de référence, et à une classe d'objets et à un objet particulier de cette classe. On peut évaluer sous cet angle le fonctionnement

logique du nom propre et des descriptions définies. Cette évaluation engage à privilégier les descriptions définies de la logique qui demandent à être spécifiées d'un point de vue linguistique. L'étude des propriétés linguistiques qui doivent caractériser l'unité utilisée comme descripteur permet de capter une différence d'« état » entre le descripteur de l'indexation et le descripteur de la recherche documentaire. Si la forme linguistique du groupe nominal « complexe » constitue l'élément privilégié pour la construction du thème de discours en recherche d'information, c'est la forme linguistique de la synapsie qui permet aux indexeurs de pouvoir manipuler, organiser, trier, choisir les accès qu'ils mettent à disposition dans un système d'information.

Dans les deux cas, le descripteur se conçoit comme une unité extraite des textes eux-mêmes. C'est, là encore, en exploitant les discours eux-mêmes et la façon dont ils permettent la construction de la référence que l'indexation peut trouver « le mode d'emploi » du descripteur et les principes de sa définition. L'indexation, comprise sous le seul angle de la détermination des descripteurs, s'apparente alors à un processus d'extraction d'unités de discours qui idéalement se réalise de façon automatisée. Le projet de l'indexation - permettre la construction de thèmes de discours - présente cette caractéristique d'être inachevé d'un point de vue interprétatif. Capter des éléments susceptibles de participer à la construction d'un thème de discours suppose en effet un type de lecture qui se situe en deçà de l'interprétation : un indexeur humain peut difficilement se livrer à tel exercice.

C'est ainsi que la reformulation de l'indexation sur la base de fondements linguistiques rencontre une faisabilité technique actuelle : on dispose aujourd'hui de systèmes automatiques d'extraction d'unités de discours. Toutefois, de tels systèmes ne sauraient être considérés comme des systèmes d'indexation automatique. Tout au plus pourraient-ils servir dans le cadre d'une indexation assistée par ordinateur. L'indexation ne se définit pas uniquement sous l'angle de l'extraction d'unités de discours ; elle consiste essentiellement en l'organisation d'un espace de discours spécifique qui, lui, engage un travail scientifique sur les sources elles-mêmes, dont on envisage difficilement, pour le moment, les conditions d'automatisation.

II.3- L'indexation dans le cadre des problématiques des « technologies de l'information »

Comme on le voit, l'indexation peut se laisser approcher sous un angle qui permet d'intégrer les « nouvelles technologies de l'information » sans s'y laisser ni réduire ni dissoudre. Il importe sur ce point d'insister sur ce qui constitue la

spécificité de l'indexation, et plus généralement, celle des bibliothèques ou des centres de documents :

On veut parfois se passer de bibliothèque, de laboratoire, de collection sans perdre ni le savoir ni la raison. C'est croire à la « nature se dévoilant aux yeux de la science ». Les chercheurs font bien autre chose que de contempler le monde dans un dérisoire *peep-show*. [...] C'est parce que les laboratoires, les bibliothèques et les collections se branchent sur un monde qui reste sans eux incompréhensible qu'il convient de les soutenir si l'on s'intéresse à la raison.
Latour, 1996, respectivement p. 43 et p. 45.

En mettant l'accent sur ce qui, en indexation, légitime le rapprochement de textes par des « mots » - le mode d'organisation des documents -, on se donne le moyen de pouvoir discuter la pertinence, en matière de recherche documentaire, des nouveaux réseaux de communication, comme le réseau Internet.

On peut en effet douter que le principe de fonctionnement du réseau Internet, tel qu'aujourd'hui il existe, soit en mesure d'améliorer, de faciliter la recherche d'information : à l'exigence documentaire d'organisation des sources s'oppose la volonté de « désenclaver » les données de leur source, de leur lieu de production, de leur temporalité. Si l'information documentaire doit trouver, sur les nouveaux réseaux de communication, une autre forme d'existence, il est nécessaire que les professionnels intègrent le débat, en mettant en avant les spécificités du traitement des documents qu'ils effectuent. Il importe sur ce point de pouvoir distinguer, dans l'indexation, une phase d'extraction des unités de discours et une phase, préalable et indispensable, d'organisation des discours.

III- De nouvelles pistes de recherche

Cette recherche a été conduite dans le but de faire émerger les propriétés linguistiques qu'exploitent les pratiques d'indexation, le plus souvent de façon implicite. Une fois circonscrits les aspects qui relèvent exclusivement du fonctionnement de la langue et du langage eux-mêmes se révèle l'étendue de ce qui constitue en propre l'indexation : ce que nous avons appelé sa stratégie d'exploration des sources et sa stratégie d'exposition des documents.

Pour spécifier ces deux aspects de l'indexation, l'approche linguistique ne suffit pas : d'autres disciplines doivent être convoquées.

Comment dégager des critères de sélection des sources qui tiennent compte de leurs conditions d'existence ? Le modèle des formations discursives proposé comme « horizon théorique » ne fournit que le principe général : prendre la mesure des « systèmes de la discoursivité ».

Comment spécifier davantage de telles règles de sélection des sources ? Sans doute doit-on se tourner vers d'autres domaines de savoir, de type sociologique et historique, voire philosophique. On rencontre alors le paradoxe suivant que les fondements de l'indexation établis du point de vue de la théorie linguistique ne tiennent que sur la base du principe de l'interdiscours, mais cet interdiscours ne peut être spécifié du seul point de vue linguistique.

De même, le point de vue linguistique montre que la stratégie d'exposition des documents doit être pensée comme telle, en dehors des langages documentaires de type classificatoire qui reposent, tout comme l'indexation, sur des présupposés dont on peut montrer les limites. Vers quel autre type d'outil se tourner ? Les procédés d'exposition des connaissances dans le discours de vulgarisation scientifique fournissent quelques pistes : comment les exploiter dans le cadre d'un « discours » documentaire qui ne se « dit » pas mais qui fonctionne plutôt comme une transformation de discours existants ?

En outre, une réflexion menée d'un point de vue linguistique engage à distinguer, dans la notion d'accès documentaire, plusieurs niveaux : l'accès à un ensemble de documents et l'accès à l'« information » ne mettent pas en jeu les mêmes procédés. On a proposé que l'accès à un ensemble de documents se fasse d'abord par le biais de « domaines ». L'accès à l'information s'établit, lui, sur la base des éléments d'une chaîne de référence. Comment articuler ces différents types d'accès ?

Le seul point de vue linguistique, s'il permet d'aboutir à de telles conclusions, ne permet pas toujours de les spécifier.

Au terme de cette recherche apparaissent une multitude de questions, laissées ici en suspens. Ces questions nécessitent d'être résolues pour que l'indexation, explicitement fondée d'un point de vue linguistique, puisse pouvoir être effectivement pratiquée. La seule mise en valeur des propriétés linguistiques à l'œuvre en indexation ne suffit pas toujours à modifier les pratiques existantes : extraire, des textes, des syntagmes nominaux ou des synapsies ne permettra pas, nécessairement, d'améliorer de façon significative l'indexation et la recherche documentaires. Il s'agit là d'une condition nécessaire mais pas suffisante.

En outre, pour pouvoir maîtriser les deux types de stratégie par lesquels nous avons proposé de concevoir l'indexation, il est indispensable de pouvoir définir de façon précise les notions de « source documentaire » et de « document » : nous avons indiqué quelques propriétés susceptibles de permettre de distinguer les objets « indexables » d'une part et de caractériser les objets indexés d'autre part. Mais là encore, il ne s'agit que d'une amorce : beaucoup reste à faire.

L'ampleur de la tâche à accomplir peut néanmoins être réduite par la prise en compte des travaux menés sur d'autres pratiques professionnelles. Sur ce point, la pratique de la terminologie et celle de la vulgarisation nous sont apparues de nature à pouvoir être profitablement rapprochées des problématiques documentaires. Cependant, là encore, il est nécessaire de pouvoir préciser le « continuum de la diffusion des connaissances » qui traverserait terminologie, indexation et vulgarisation scientifique : où placer l'indexation ? Comment se singularise son mode de diffusion de la connaissance ?

Les questions à résoudre qui permettraient à nos propositions de trouver une voie de réalisation effective sont nombreuses et importantes. Reste que l'approche de l'indexation proposée dans cette étude montre qu'il est possible de spécifier l'indexation dans le rapport particulier qu'elle entretient avec les textes eux-mêmes et les unités lexicales qui les constituent : cette tentative de « réconciliation » entre la langue et la pratique professionnelle de l'indexation nous semble en cela constituer une hypothèse de travail constructive.