

LECTURES COMPLÉMENTAIRES

SUR L'ACCÈS PAR SUJET, THÈME OU ACTIVITÉS ET L'INDEXATION

LECTURE COMPLÉMENTAIRE 1

Bureau canadien des archivistes (1992). *L'indexation par sujet en archivistique*. Ottawa. p.92-99.

Ce texte porte sur le processus d'indexation. Il présente d'abord l'objet et la finalité de l'indexation. Il expose les différentes étapes de l'indexation ainsi que les niveaux d'indexation et de description.

4.2 Indexation sujet

4.2.1 Le processus d'indexation

Objet et finalité de l'indexation

Le résumé, très utile pour condenser l'information, ne fournit pas immédiatement les clés d'accès nécessaires au repéage. Cette fonction revient à l'indexation qui consiste à choisir, à partir du résumé, s'il existe, ou bien à partir du document original, les termes plus ou moins normalisés représentant les concepts importants. La transcription en langage documentaire se fait grâce à des outils d'indexation tels que thésaurus, liste de vedettes-matière, système de classification, etc. C'est à partir des résultats de l'indexation que sont constitués les index par lesquels s'effectue le repéage.

Les étapes de l'indexation

L'indexation comprend les étapes que l'on retrouve dans toute opération d'analyse :

¹⁰² Françoise Blamoutier. «L'entraînement à la rédaction de textes documentaires en langage contrôlé destinés à la traduction automatique.» In: Troisième congrès européen sur les systèmes et réseaux documentaires. *Franchir la barrière linguistique* (Luxembourg, 3-6 mai 1977), 515.

LA CONDENSATION ET L'INDEXATION

prise de connaissance du contenu, puis identification et sélection des notions principales;
- choix des termes d'indexation généralement à l'aide d'un vocabulaire contrôlé (ISO 5963-1985)

L'organisation des outils de recherche documentaire constitue l'étape suivante qui est, toutefois, omise dans les normes. Dans le sixième chapitre, nous verrons comment établir un index imprimé.

Pour le choix des concepts, la norme préconise le recours à une grille d'indexation. Il ne suffit pas de se demander: «De quoi traite le document?», car bien souvent, un document ne raconte qu'un événement. La grille sert alors à rechercher quelle est l'action, qui a accompli l'action (agent de l'action), sur quoi a porté l'action (objet de l'action), à quel moment et en quel lieu s'est produite l'action.

Parfois les questions auxquelles le document peut répondre correspondent à d'autres catégories d'information comme par exemple : type de médicament, composition chimique, maladie traitée, posologie, contre-indications, etc. L'analyse de documents photographiques en archéologie se pratique selon la grille suivante dans un centre de documentation¹⁰³ : (1) objet du document (par exemple : statuette); 2) matière (par exemple : argile cuite); 3) iconographie (par exemple : femme); 4) origine géographique (par exemple : Anatolie occidentale); 5) datation (par exemple : VI^e millénaire av. J.-C.). La grille d'indexation est établie d'après la structure d'information commune qui se dégage de l'examen des documents de même type et d'après le genre de questions posées par les utilisateurs.

Pour l'étape de *transcription en vocabulaire contrôlé*, l'archiviste doit identifier les formes préférentielles retenues pour exprimer un concept — l'emploi de synonymes compliquerait le repéage et nuirait au taux de rappel — et accepter de nouveaux termes pour traduire de

¹⁰³

Serge Cleuziou, «Projet de code pour l'analyse d'une collection de documents photographiques en archéologie.» In: *Banques de données archéologiques*. Marseille, 12-14 juin 1972 (Paris: CNRS, 1974).

LA CONDENSATION ET L'INDEXATION

nouvelles notions, si le langage documentaire le permet, en prenant toutes les précautions nécessaires pour préserver la cohérence interne de l'outil d'indexation. La qualité du langage documentaire est extrêmement importante pour assurer la fidélité de la représentation. Comme nous l'avons déjà dit, ce langage doit comporter non seulement un vocabulaire mais aussi une syntaxe, tout comme la langue naturelle. C'est d'ailleurs cette syntaxe qui va permettre de préciser le rôle de chaque terme dans la représentation d'un événement ou encore des différentes facettes sous lesquelles un objet est analysé.

La nécessité d'une syntaxe documentaire varie selon le domaine. Les sciences pures — physique, chimie, biologie, etc. — sont des champs de la science dans lesquels les concepts étudiés sont dits «purs». Dans ces domaines, la signification précise des termes est généralement bien comprise à l'intérieur de la discipline. De plus, le mot qui traduit le concept dans ces champs aura toujours le même sens dans tous les documents indexés. Par exemple, on comprend clairement la signification des termes *piston* et *moteur*. Ces deux termes sont reliés comme partie d'un tout, le tout étant le *moteur* et le *piston* une de ses parties. Il est peu probable que ces deux termes soient reliés d'une autre façon.

Les sujets reliés aux sciences humaines, aux sciences sociales et aux arts sont à l'opposé. Les termes utilisés pour décrire une chose ou un phénomène particulier à ces domaines changent fréquemment, dépendant du point de vue de l'observateur. La signification d'un terme est variable à l'intérieur d'un champ disciplinaire. Il peut même changer — ne serait-ce que de façon minime — d'un groupe de spécialistes à un autre. Le fait que les relations entre ces concepts ne sont pas évidentes est encore plus problématique. Personne ne peut comprendre le rôle d'un concept tant que le contexte, le cadre ou la référence appropriés n'ont été établis.

Par exemple, il n'existe pas de définition largement acceptée de termes tels que *famille*, *société* ou *démocratie*. Ces termes et le terme *moteur* sont très différents et cette différence vient du fait que l'un est relié aux sciences pures et l'autre aux sciences humaines et sociales. Les rôles ou les relations ne sont pas évidentes non plus entre les termes *histoire*, *imprimés* et *bibliographie* par exemple. Ces trois

termes se retrouvent cependant tous dans les titres suivants : «L'histoire des bibliographies imprimées» | "Bibliographies de l'histoire de l'imprimé» ou "Bibliographies imprimées de l'histoire» Que recherche-t-on vraiment? Il est impossible de le savoir sans voir les termes associés dans un contexte particulier ou liés ensemble par une syntaxe particulière comme ils apparaissent plus haut.

Tel que nous l'avons relevé dans le premier chapitre, les utilisateurs d'archives (près de 64 %) sont en majeure partie diplômés en sciences humaines et sociales et 20 % seulement le sont dans le domaine de l'histoire. Il semble probable que la plus grande partie des besoins d'information des usagers soit reliée aux disciplines dans lesquelles se retrouvent l'intérêt et l'expertise de ces derniers. Ce qui est frappant, d'après ces données, c'est qu'il n'est fait aucune mention des demandes ou même des utilisateurs provenant des sciences pures.

Compte tenu du fait que les utilisateurs d'archives proviennent principalement du domaine des sciences humaines et sociales, cela a pour conséquence qu'un système d'indexation sujet pour les documents d'archives doit être plus sophistiqué qu'un système appliqué aux sciences pures. Le véritable test pour un système d'indexation est qu'il puisse fonctionner aussi dans le domaine des sciences humaines et sociales. Dans ces sciences, il ne suffit pas d'avoir un index composé simplement d'une liste des termes ou des mots clés. En plus des termes, on doit trouver un moyen de démontrer comment ils sont reliés les uns aux autres dans le contexte des documents indexés. En d'autres mots, il doit y avoir à la fois des termes et de la syntaxe. S'il n'en est pas ainsi, un grand nombre de documents non pertinents seront inévitablement repérés. S'il est impossible de spécifier, d'une façon ou d'une autre, une relation syntaxique entre *albums* et *photographies* — comme *album de photographies* —, on retrouvera en réponse tous les documents non pertinents à cette recherche qui portent, par exemple, sur des *photographies d'albums*.

Il a déjà été admis que le processus d'indexation sujet comporte deux étapes intellectuelles distinctes à savoir l'analyse conceptuelle du sujet du document ou de l'ensemble des documents et la traduction de cette analyse en un vocabulaire particulier à un système d'indexation. Les deux étapes sont reliées!

Le degré de sophistication du vocabulaire utilisé par l'indexeur à l'étape de traduction a un impact considérable sur le degré et le contenu de l'analyse conceptuelle. Ce serait une perte de temps pour un indexeur d'analyser le contenu d'une pièce avec grande minutie et en utilisant des termes très spécifiques, si le vocabulaire disponible pour la traduction consistait seulement en un nombre limité de termes plutôt généraux. Par exemple, si le vocabulaire des indexeurs est limité à l'utilisation du terme *animaux* pour un dossier sur les licomes, chevaux et zèbres il serait vain d'analyser ce dossier en dessous du niveau général.

Un autre problème de traduction apparaît pour les indexeurs travaillant dans les archives ou les musées lorsque les fonds et les collections à indexer sont constitués de documents non textuels. L'analyse conceptuelle est d'autant plus importante et la traduction de cette analyse encore plus problématique lorsque l'indexeur doit exprimer à nouveau en mots le contenu de photographies, diapositives et autres pièces dont le langage est visuel, structurel, etc. Le cinquième chapitre porte entièrement sur ces langages documentaires.

Niveaux d'indexation et niveaux de description

Ce que nous avons dit des liens entre niveau d'analyse et niveau de description vaut, bien sûr, pour l'indexation. En archivistique, à cause du volume de documents, on commencera par indexer au niveau le plus général (le fonds, la série, etc.) pour finir par l'indexation de pièces ayant une grande valeur. Dans le cas de certains médias, on peut même analyser à un niveau encore plus fin: une séquence de film, certaines photographies dans un montage.

L'indexation d'un document ou d'un ensemble de documents d'archives peut se faire à différents niveaux. On peut donc parler d'indexation à un macro-niveau et d'indexation à un micro-niveau. Au macro-niveau, l'indexation reflète le document comme un tout, et est, en fait, un résumé sujet du contenu. Un film peut être réalisé, par exemple, sur le ballet au Canada. L'indexation de cette œuvre au macro-niveau ou résumé devra fournir un accès sujet sous les concepts généraux de *ballet* et de *Canada*. Cependant, si l'indexeur analyse le même film segment par segment, il est évident que certains sujets

LA CONDENSATION ET L'INDEXATION

spécifiques pourront être révélés. Ainsi, le film peut inclure une séquence de Karen Kain. Il peut aussi inclure la tournée du Royal Winnipeg Ballet dans les Maritimes. Indexer à un micro-niveau ou à un niveau analytique exigera de fournir des accès sujets sous certains termes tels : *Karen Kain, Royal Winnipeg Ballet* et probablement d'autres termes à propos de la tournée dans les Maritimes (avec les dates).

Un autre exemple sera illustré par une brochure publiée par un département de santé et d'hygiène des animaux domestiques. La brochure contient un certain nombre de pages relatives aux chiens, une ou deux pages relatives aux chats, et quelques pages sur les perruches. Si cette brochure est indexée à un macro-niveau, on fournira un accès sujet sous le terme général *animaux domestiques*. Par ailleurs, au micro-niveau, les termes indexés seront *chiens, chats et perruches*. Somme toute, il existe une relation entre les niveaux d'indexation, les niveaux de spécificités et les niveaux d'exhaustivité. Il existe aussi une relation entre les niveaux d'indexation et les niveaux de description archivistique. Dans la majorité des cas, il serait illogique et inefficace d'appliquer l'indexation de micro-niveau à la description du fonds. De même, l'indexation de macro-niveau devrait permettre de fournir des accès sujets spécifiques à *chacun* des niveaux archivistiques : à la série, au niveau de la série, au dossier, au niveau du dossier et à la pièce au niveau de la pièce. Au niveau de la pièce cependant, l'indexation de type *macro* ne doit pas toujours être considérée suffisante comme dans le cas du film et de la brochure ci-haut mentionnés.

L'essentiel est de faire en sorte que, dans l'index, la structure de la référence soit représentative du niveau de description correspondant au matériel analysé. Comme il a été mentionné dans le deuxième chapitre, les sujets doivent être clairement reliés à un niveau de description (fonds, série, dossier, pièce à indexer) et doivent également montrer la relation hiérarchique selon la description à plusieurs niveaux.

Le choix du niveau d'indexation dépend des besoins et des ressources. Il faut plus de temps et plus d'expertise pour analyser des documents individuels plutôt qu'un ensemble de documents. Le résultat est évidemment plus volumineux et par conséquent plus coûteux.

Étant donné les habitudes de description et d'analyse dans les services d'archives, mieux vaut commencer par indexer non pas les documents eux-mêmes, mais les représentations secondaires à savoir les annotations ou les résumés qu'en offrent les instruments de recherche.

Il reste encore beaucoup à apprendre sur l'indexation. La prise de conscience de la complexité de cette opération, comme de l'analyse en général, s'est faite au fur et à mesure du besoin d'automatisation de l'ensemble du processus, de ses étapes, de même que de la production des outils et des produits :

[...] indexing is an extremely complex activity, inextricable from the classification - that is, structure - of the field of discourse concerned, the verbal expression of its structure, methods for presenting the index and rules for its use to readers, and methods and languages and codes for matching it to the retrieval processes¹⁰⁴.

Génie constate avec un certain pessimisme :

Information retrieval from bibliographic databases is relatively crude and full of uncertainty, compromise, and approximation. Essentially, it is not suited to provide sophisticated document representation¹⁰⁵.

Heureusement, les politiques et grilles d'indexation pallient, en partie, à l'arbitraire et au manque d'uniformité. De plus, certains langages documentaires permettent de rendre plus fidèlement que d'autres le contenu des documents.

¹⁰⁴ R. Fairthorne, «Content Analysis, Specification and Control.» *Annual Review of Information Science and Technology* 4 (1969): 77.

¹⁰⁵ Brenda Genie, *Online Information Systems: Use and Operation ...*, pp. 170-171.