



68th IFLA Council and General Conference

August 18-24, 2002

Code Number:	097-144-F
Division Number:	IV
Professional Group:	Classification and Indexing with Cataloguing - <u>Workshop</u>
Joint Meeting with:	CILIP Cataloguing and Indexing Group
Meeting Number:	144
Simultaneous Interpretation:	-

Bibliothèques et archives: intégration de professionnalisme dans le thésaurus du cadre de classement des Universités italiennes

Giovanna Granata, Gianni Penzo Doria & Zanetta Pistelli

Universita degli Studi di Padova
Padua, Italy

L'attention que les archivistes ont consacré aux archives courantes n'est pas encore comparable à celle adressée aux archives historiques. Le papier de tournesol est représenté non tant de l'insuffisance des études archivistiques sur le cadre de classement, mais plutôt de l'absence de critères rédactionnels.

Le cadre de classement, vrai acte d'organisation pour un sujet producteur, doit être rédigé selon des critères scientifiques, en évitant de le réduire à une banale énumération de bureaux ou d'organismes qui échangent de la correspondance administrative, et surtout en tenant bien distincte l'organisation logique de l'organisation physique, non seulement des documents, mais aussi de l'organisme producteur même. Il faut aussi tenir compte que le cadre de classement est un instrument technique, qui, dans les mains d'opérateurs non qualifiés, peut provoquer des conflits et des incertitudes sur la cohérence de son application.

Afin d'arriver à une certaine normalisation, les universités italiennes associées au projet "*Titulus 97*" (http://www.unipd.it/ammi/archivio/tit_idx.htm) ont commencé depuis 1997 à élaborer un cadre de classement commun, accompagné d'un index, afin d'obtenir un instrument d'orientation souple et efficace dans le schéma de classement : les "dénominations" identifiées à partir des documents ont été ordonnées d'après le premier terme et manipulées à travers la rotation des termes plus significatifs. Cette rotation, d'une part a augmenté les possibilités d'accès, mais de l'autre a

souvent comporté une perte d'intelligibilité. On a eu immédiatement la perception qu'il était nécessaire de construire des énoncés moins ambigus et plus contrôlés et pourtant nous sommes convaincus que la collaboration entre les deux professionalismismes des archivistes et des bibliothécaires était une stratégie gagnante, tout en conservant la correcte distinction des rôles et des méthodes.

Nous avons donc travaillé tant sur le versant de la syntaxe que sur celui de la terminologie, en affrontant plusieurs problèmes, et avant tout celui de la variabilité des constructions employées dans les "dénominations". En effet, les "dénominations" sur lesquelles nous sommes intervenus présentaient une articulation particulière, typique du langage bureaucratique, plein de sous-entendus et de locutions "argotiques" pour dénoter les procédures et les actes. Ce langage est déjà, en soi, très compliqué, puisque il se réfère à des objets complexes (les normes, l'organisation, les ressources, les rapports extérieurs) à l'intérieur d'une organisation complexe. De plus, il ne sert pas à décrire du matériel bibliographique, mais des documents administratifs, et la différence est notable. Si il est vrai que l'objet dont, selon la règle, le document même porte information peut être comparable au titre d'un document bibliographique, il faut toutefois admettre que dans ce dernier cas on peut compter sur une coutume bibliographique vieille de plusieurs siècles, selon laquelle, sauf en cas de titres obscurs imaginaires ou encore ouvertement déviants, le titre contient déjà, en forme sommaire, le thème d'ensemble du document. Au contraire, l'objet des documents administratifs, loin d'avoir une formalisation et standardisation suffisante, s'exprime souvent par des références précises et circonstanciées, avec des détails à grappe ou en emboîtement autour d'un noyau central, d'une manière analogue avec ce qu'il se produit dans nos textes de lois, où la syntaxe présente une succession continue de subordonnées.

Ainsi, on a dû reformuler les "dénominations" et en transformer les énoncés en véritables "condensés" d'indexation, avec un ordre de citation rigoureux et cohérent et un sévère contrôle terminologique.

D'une façon complètement opposée à la complexité de quelques « condensés », un autre problème s'est présenté. Dérivant de l'index du cadre de classification, les « dénominations » privilégient dans différents cas une certaine généralité par rapport à la spécificité des « condensés » propres au sujet. Il s'agit de nombreux mots - par exemple "activité de représentation" ou "règlements" - constitués seulement de termes substantivés, qui comptent évidemment grouper ensemble les documents relatifs aux différentes situations connexes, indépendamment de la signification de chacune. Dans ces cas, le contexte de référence nous a été utile, constitué du titre et de la classe convenable. Par exemple, pour des "dénominations" telles que "cadres de tableaux" ou "dommages", seules les références au titre et à la classe relative permettent de comprendre qu'il s'agit, pour le premier, de l'acquisition et de la fourniture de cadres de tableaux et pour le second de contentieux à la suite de dommages causés aux personnes ou aux biens de l'Université.

Une fois le sujet reconstitué, il fallait l'exprimer.

Selon quelles syntaxes et quels principes?

La référence plus immédiate ne pouvait être que le "Soggettario" de la Bibliothèque Nationale Centrale de Florence, lequel, toutefois, a été soumis à plusieurs critiques, dont la principale est qu'il s'agit d'un outil typique de l'époque du papier, dans lequel l'efficacité de la récupération de l'information est remise à la significativité du mot d'ordre. Celui-ci occupe la première position

dans le « condensé » et impose un ordre de citation fondé sur l'importance présumée des termes auxquels assurer un accès direct. Cela rend difficile l'expression des sujets présentant des notions différentes combinées ensemble, pour lesquels, pourtant, il devient nécessaire de mettre en place plus d'un "condensé", sachant qu'aucun n'est réellement apte à couvrir tout le champ sémantique du sujet.

Notre choix s'est donc porté sur une méthode d'indexation élaborée par le GRIS (Gruppo di Ricerca sull'Indicizzazione per Soggetto), méthode qui non seulement permet d'exprimer des "condensés" aptes à couvrir tout le champ sémantique du sujet, mais encore en assure l'ordre de citation selon une cohérente et rigoureuse intelligibilité. Dans son élaboration le GRIS remonte aux études sur les fondements théoriques de l'analyse par sujet et les classifications développés surtout en Angleterre par le CGR (Classification Research Group) et aboutit à la formalisation du système PRECIS. Dans la Bibliothèque Nationale Centrale de Florence un projet de révision du "Soggettario" à la lumière de la méthodologie du GRIS est en cours.

En suivant les normes du guide GRIS¹, auquel nous renvoyons pour une présentation plus analytique des objectifs et de la méthodologie, nous avons donc procédé à identifier dans les "dénominations" le rôle de chaque notion: action, objet, agent etc., et nous avons ensuite disposé les termes du "condensé" selon l'ordre de citation assigné à chaque rôle et selon le critère de la dépendance du contexte. Ci dessous nous donnons un exemple pratique de la méthodologie suivie. Considérons un énoncé de sujet, présent dans l'index au cadre de classement, tel que: "Mod. 101, prédisposition, personnel enseignant". Il s'agit d'un "condensé" apparemment simple, dans laquelle on peut reconnaître une action (prédisposition), un objet (mod. 101), un bénéficiaire (personnel enseignant). En réalité, la notion "mod. 101" n'est pas claire; il s'agit d'un formulaire nécessaire pour la déclaration des revenus du corps enseignant; donc le "condensé" doit être recomposé ainsi: prédisposition du formulaire 101 pour la déclaration des revenus du corps enseignant. A partir de là, on peut identifier deux noyaux conceptuels, à leur tour décomposables en notions ultérieures: le premier est "prédisposition du formulaire 101", le second "déclaration des revenus du corps enseignant". Sur la base de l'analyse des rôles, l'ordre de citation sera le suivant:

Corps enseignant – Revenus - Déclarations – Formulaires : Mod. 101 – Prédisposition

Par rapport à la "dénomination" d'origine les avantages obtenus sont la meilleure intelligibilité des notions, exprimées dans l'index au cadre de classement en forme trop synthétique et argotique; la restitution des rôles de la sémantique et de la syntaxe; l'accès à tous les termes sans rotation ni inversion.

Nous nous sommes également posé la question, surtout dans le cas d'énoncés plus complexes, de la rigide analyticit  dans l'identification des rôles et dans l'expression des termes équivalents qui a déterminé une parcellisation plus grande des "condensés" par rapport aux habitudes du langage naturel. Par exemple:

Doctorats de recherche - Concours - Jurys d'examen - Nomination - Communication aux coordonnateurs de doctorat

¹ Associazione italiana biblioteche. GRIS-Gruppo di ricerca sull'indicizzazione per soggetto. *Guida all'indicizzazione per soggetto*. - Roma : A.I.B., 1997.

Sans doute les "condensés" ainsi reconstitués, par rapport à un instrument déjà préparé et maîtrisé par les collègues archivistes, peuvent poser plusieurs problèmes quant aux avantages pratiques et à la facilité pendant l'application: le langage naturel, même si soumis à des contorsions lexicales pour permettre une récupération de l'information plus probable, est beaucoup plus immédiat qu'un langage contrôlé et très formalisé.

Toutefois, l'adoption d'une syntaxe cohérente pour la construction et l'intelligibilité même des "condensés" nous a semblé un avantage trop important. En effet, si le cadre de classement, et donc son index de référence, est destiné à devenir un outil valide pour toutes les universités italiennes qui ont adhéré au projet "Titulus 97", alors des instruments très standardisés, contrôlables, ajournables selon des critères transparents et prévus à cet effet sont nécessaires pour une bonne coopération.

A côté du travail de construction des sujets on a du affronter celui du contrôle terminologique: à quel vocabulaire se référer? En effet dans les "dénominations" utilisées comme index du cadre de classement, malgré la connotation juridique-administrative du lexique, le manque d'éléments de contrôle avait donné lieu à la prolifération d'une variété de synonymes ou de presque-synonymes, souvent utilisés de manière impropre, par exemple personnel, dépendants, travailleurs, etc. Parfois la compréhension exacte du sens des "dénominations" était handicapée par l'usage d'homonymes, par exemple "diplômes universitaires" qui en Italie indique à la fois une typologie de cours d'étude de niveau plus bas que la maîtrise, mais aussi les attestations qui certifient la conclusion d'un cours d'étude: diplôme de maîtrise, diplôme de doctorat. Finalement un autre problème venait de l'usage d'expressions composées, typiques du langage de la bureaucratie, par exemple "périodes de service ouvrables près d'organismes publics" ou "payements au contentieux" qui proposaient encore une fois au niveau terminologique les difficultés du contrôle syntaxique: de ce point de vue s'est posée l'exigence du choix de termes plus simples et de les coordonner au moment de la formulation des sujets.

En raison de la complexité des problèmes, la définition du vocabulaire ne pouvait pas aboutir tout simplement à une liste plate, mais on a reconnu, dès le début, la nécessité d'employer un thesaurus pour structurer les termes en respectant toutes les interrelations réciproques. Le contexte coopératif de Titulus 97 a porté un autre appui à cette décision. D'une part la coopération d'organismes différents rend encore plus nécessaire de disposer d'instruments raffinés et précis de contrôle terminologique; de l'autre elle peut garantir un support valide à la nécessité de la mise à jour qui constitue une des principales difficultés inhérentes au choix théssaural.

Pour la construction du thesaurus on a évalué la possibilité de procéder d'en bas, c'est-à-dire de la terminologie même, ou de partir d'instruments existants. À ce propos, quelques réalisations bien connues, nées en milieu éducatif (Eurovoc) ou en milieu juridique-législatif (TESEO) ont été examinées. Cependant les choix de caractère disciplinaire ne se sont pas révélés fonctionnels aussi bien en ce qui concerne les champs sémantiques couverts, que pour le niveau de spécificité demandé. Un modèle plus fonctionnel est venu du TRT (Thesaurus régional toscan) qui est construit sur la base d'une catégorisation par facettes de type général et abstrait, en accord avec les critères suggérés par le même GRIS qui nous a guidé dans les problématiques du contrôle syntaxique.

Selon de tels critères, le regroupement des descripteurs en facettes a été effectué par TRT sur la base de la nature même des idées qu'ils représentent, par exemple activité, instruments etc., indépendamment de leur contexte spécifique d'utilisation. Par conséquent il offre une modalité d'organisation la plus objective possible qui s'adapte le mieux à la complexité de la terminologie et en permet aussi plus aisément la mise à jour. En outre, bien qu'il ait été développé en milieu bibliothécaire, il est né dans un contexte qui est voisin de celui d'où vient notre terminologie: la bibliothèque de type juridique-administratif qui est le support des activités de l'organisme « Région Toscane ». C'est dans ce but que les facettes abstraitement suggérées par GRIS ont été déterminées plus analytiquement afin de permettre le traitement de la terminologie en question. Ces dernières sont: Actes juridiques, Activité, Conditions, Disciplines, Formes, Objets, Organisations, Gens et groupes, Procès, Place, Instruments, Structures, Temps, Théories et mouvements.

Une fois effectuée le choix de l'ossature générale de TRT comme référence structurale on a du en vérifier l'hospitalité par rapport à la terminologie des "dénominations."

À ce sujet on a tenté avant tout de déterminer les descripteurs, en utilisant comme critère guide pour les problèmes lexicaux, au-delà des choix déjà effectués par TRT, la règle ISO 2788/1986. Les principes de la règle ont été appliqués avec rigueur, tout en tenant compte des problématique inhérentes à la typologie du langage bureaucratique; en particulier, en ce qui concerne les termes composés qui, comme on a vu, abondaient dans l'index du cadre de classement, on a parfois choisi de les admettre comme "acceptés" sans ultérieurement les réduire aux expressions plus simples, pour ne pas perdre leur spécificité.

Dans le choix entre singulier et pluriel on a adopté le critère de la dénombrabilité, en conformité avec les indications offertes par le guide GRIS et adoptées par TRT: on a utilisé donc le singulier pour les termes qui indiquent une activité et le pluriel pour ceux qui sont utilisés en forme nominale; dans quelques cas, cependant, le glissement sémantique connexe à l'usage du singulier et du pluriel a suggéré de maintenir les deux formes, le singulier pour l'action, le pluriel pour les résultats de l'action même: actes, documents etc.

Pour le traitement de noms propres, c'est à dire d'organismes, fondations, consortiums, projets, répartitions administratives, utilisés dans les dénominations on a opté pour la construction d'une liste d'autorité extérieure au thesaurus, plutôt qu'inclure chacun dans la hiérarchie du thesaurus même, étant donnée l'extrême variété des descripteurs pour les universités du projet Titulus 97. On a par contre inséré dans la hiérarchie théaurale le terme "commun": chaque université a, en effet, pour son centre de calcul une dénomination différente, qui toutefois se rapporte à la même réalité fonctionnelle.

L'insertion des termes dans les facettes de TRT a donné de bons résultats en montrant l'hospitalité de sa structure, même si l'expansion de la terminologie concernant les aspects de l'organisation universitaire par rapport au reste des descripteurs a entraîné quelques problèmes étant donnée l'extrême spécificité des articulations et des facettes "Gens et groupes", "Organisations", "Activité". Ce manque d'homogénéité ne met cependant pas en discussion la structure générale du thesaurus, de façon qu'on peut supposer le développement du dictionnaire grâce à l'ajournement de la terminologie.

Pour la gestion des relations thésaurales nous avons eu recours à un programme dénommé "Beat". Réalisé par Josep Sau du Centre d'informatique l' « Universitat de Barcelona », il est disponible gratuitement pour usage non commercial à l'adresse Internet <<<http://www.willpower.demon.co.uk.thessoft.htm#BEAT>>> d'où il est possible de le télécharger. "Beat" soutient la gestion de toutes les relations prévues par la règle ISO mentionnée ci-dessus: la relation d'équivalence, qui renvoie du terme rejeté, mais conservé dans le thesaurus comme clé d'accès, à celui accepté; la relation hiérarchique, qui réunit chaque terme accepté avec celui sovra-ordonné dans la structure du thesaurus; l'associative, qui signale pour chaque terme ceux qui sont corrélés dans le contexte sémantique couvert par le thesaurus.

Finalement, pour ce qui concerne les résultats, "Beat" fournis plusieurs types de sorties, systématique, alphabétique, permutée.

Pour l'emploi du thesaurus, au-delà à l'output hiérarchique, indispensable pour reconstruire les relations thésaurales, il nous est paru nécessaire d'offrir une sortie de type alphabétique dans la quelle chaque terme est indiqué avec l'ensemble de ses relations.

En plus, pour éviter la rotation des termes dans les sujets, préparée par les collègues archivistes en phase de construction premier de l'index des "dénominations", nous avons associé chaque terme dans la liste alphabétique du thesaurus aux sujets dans lesquels il paraît, de façon à consentir aux utilisateurs d'accéder à chaque sujet, exprimé dans son caractère exhaustif et parfaitement intelligible, à partir d'un quelconque de ses éléments.

Pour garantir la visibilité réciproque de termes et sujets, un programme ad hoc à été réalisé qui complète les uns et les autres.

En dernière analyse, pour chaque descripteur de l'output alphabétique, les sujets dans lesquels il est employé sont aussi indiqués, au-delà de l'ensemble de ses relations hiérarchiques. Les sujets portent la référence au titre et à la classe dans laquelle ils doivent être inclus, de façon que thesaurus et index du cadre de classement ne restent pas indépendants, mais qu'il soit possible de passer de l'un à l'autre en contrôlant aussi bien la syntaxe que la sémantique.

Ceci permet aux opérateurs, d'un côté d'avoir un guide à l'introduction de nouveaux sujets par analogie avec ceux qui existent déjà, de l'autre de ne pas les reproduire en effectuant le contrôle à partir de descripteurs pertinents.

.....