

# Le Thésaurus Eurovoc de l'Union européenne

## 11 langues pour 15 pays

MANUELA MASCARENHAS PINTO

La troisième édition du thésaurus Eurovoc de l'Office des publications officielles des Communautés européennes et du Parlement européen est disponible sur papier et sur CD-ROM (environ 12 000 descripteurs et non-des-cripteurs par langue). Eurovoc, vocabulaire européen, est le langage d'indexation de plusieurs bases de données de diffusion d'information.

Eurovoc se présente renouvelé et actualisé, il a atteint le but qui lui a été assigné: l'indexation de la production bibliographique communautaire post Maastricht. Il a été créé en 1984 pour l'indexation des publications des Communautés européennes, de leurs actes officiels publiés au Journal officiel et de la documentation du Parlement européen.

Son atout premier est le multilinguisme (les onze langues de la Communauté européenne) qui le rend unique dans le panorama documentaire international. L'indexeur peut choisir la langue d'indexation parmi un large éventail d'options et celle-ci peut ne pas coïncider avec la langue de la recherche ultérieure.

En 1984 sortait sa première édition en sept langues et en 1987 la deuxième édition, en neuf langues. Cette troisième édition est traduite en onze langues, avec la nouveauté du suédois et du finnois (en phase finale de traduction). Eurovoc est devenu un instrument d'indexation pour les quinze pays de l'Union européenne.

Eurovoc est de plus en plus reconnu comme l'outil documentaire de la bibliographie communautaire, utilisé par des parlements nationaux des États membres, des centres de documentation privés et plus récemment encore l'intérêt manifesté par les pays scandinaves et de l'Europe centrale et orientale est indéniable. Certains de ces derniers pays l'ont adapté à leurs besoins nationaux et en font des traductions non officielles.

## La logique interne d'Eurovoc

Ce langage contrôlé se base sur quatre piliers:

1. Post-coordination d'unitermes (existence de quelques descripteurs composés, comme exception).
2. Structuration hiérarchique des descripteurs en microthésaurus, placés sous de grands domaines thématiques.
3. Non-ambiguïté des descripteurs: rejet de la synonymie par le choix d'une seule forme au détriment des termes de proximité inscrits comme non-descripteurs; rejet de l'homonymie par l'emplacement de chaque descripteur dans un domaine thématique précis; rejet de la polyhiérarchie par l'inclusion d'un descripteur uniquement dans un seul domaine et dans un seul champ sémantique, rejetant d'autres domaines également possibles (toutefois, Eurovoc admet exceptionnellement la polyhiérarchie dans le domaine *géographique et organisations internationales*).
4. Respect par la *richesse lexicologique* de chaque langue démontrée par les non-descripteurs qui varient de langue en langue et pour chaque langue, d'un domaine à l'autre.

## Trois présentations

Eurovoc se présente en trois volumes: présentation alphabétique permutée, présentation thématique et présentation multilingue.

L'indexeur dans la tâche d'indexation/recherche bibliographique peut avoir deux raisonnements différents: soit faire la recherche par mot soit par thème.

S'il veut rechercher par mot, il se sert de la présentation alphabétique permutée. Sa structure du type promptuaire/dictionnaire, le découpage et la rotation des mots significatifs des descripteurs est une aide précieuse pour donner une idée des descripteurs possibles. Dans la présentation alphabétique permutée le poids se pose sur les descripteurs: le tri est fait par descripteur, chaque descripteur présente son environnement sémantique, termes associés, la chaîne de ses non-descripteurs, et éventuellement une *scope note* pour bien le préciser. La présentation alphabétique aide beaucoup à cerner le sens de chaque descripteur.

S'il a une position différente, s'il veut chercher par sujet, il doit travailler davantage avec la présentation thématique. Celle-ci est structurée thématiquement et suggère à l'indexeur d'autres thèmes également pertinents. Elle est fondamentale pour la vision globale des microthésaurus car elle offre un ensemble très étendu de thèmes organisés sous le sujet du microthésaurus. Les relations associatives des descripteurs y sont affichées.

L'idéal pour l'indexeur sera de combiner les deux approches, ce qui est plus souvent le cas. La présentation multilingue donne la traduction des descripteurs dans les différentes langues communautaires et montre la richesse linguistique d'Eurovoc. Cette présentation est devenue un outil de référence terminologique.

### **Mise à jour**

Toute actualisation et mise en forme est une opportunité pour une révision technique, linguistique et terminologique des différentes versions linguistiques. Ces actualisations périodiques sont louables, car elles sont sérieuses, profondes et bien adaptées à la logique interne d'Eurovoc. Ainsi, Eurovoc s'est régénéré selon les besoins de ses utilisateurs, lesquels maintiennent avec l'*Unité de maintenance* un dialogue permanent, vif et créatif.

Cette troisième édition n'a pas seulement apporté de nouveaux descripteurs et non-descripteurs, elle a aussi amélioré les termes existants et leurs relations, pour répondre aux demandes des indexeurs et aux observations de l'équipe linguistique. L'*Unité de maintenance* a décidé peu de suppressions, seulement quelques descripteurs faisant double emploi ou dont le sens était ambigu. Descripteurs avec un poids historique comme «guerre froide» ont continués.

Plus le nombre de relations sémantiques par descripteur est élevé plus un thésaurus est difficile à mettre à jour. Mais, par contre, un thésaurus bien structuré, comme Eurovoc, facilite les ajouts, car le sens de chaque descripteur est univoque.

Pour prendre une décision adéquate pour chaque terme on a procédé à l'analyse de la fréquence d'utilisation des descripteurs à l'indexation et à l'interrogation: si un terme était employé souvent à l'indexation et presque jamais à la recherche ou bien il était trop spécifique ou le contraire. Il a fallu l'adapter au type de recherche:

l'élargir car la recherche ne se faisait pas à un niveau si profond ou bien le subdiviser en termes plus spécifiques.

C'est la fréquence de la recherche qui doit décider de la position d'un descripteur dans le thésaurus. Une position plus importante s'il est très utilisé, ou une position plus discrète s'il est moins demandé. L'utilisation peut même faire inverser un descripteur en non-descripteur et vice versa.

L'emploi d'Eurovoc a fait naître une liste de candidats descripteurs qu'il a fallu analyser afin de juger de son intégration. Des statistiques d'utilisation ont été faites pour pouvoir juger de chaque candidature. Le degré de spécificité de chaque candidat fut évalué, trop pointu ou trop large. La possibilité de combiner des descripteurs existants fut avancée, pour exprimer le concept proposé. Il a fallu vérifier si un synonyme d'un candidat était déjà présent dans le thésaurus ou avec le statut de non-descripteur.

Et avant de faire une addition, une suppression ou modification de relation, il a été veillé à ne pas bouleverser la cohérence interne du thésaurus car tout ajout/modification de descripteurs et de ses relations a un impact sur les relations existantes et peut même obliger au changement de la structure du thésaurus.

En cas d'une suppression ou rétrogradation d'un descripteur au rang de non-descripteur, il a fallu réindexer en utilisant le terme directement supérieur et faire attention à l'équilibre de la structure globale perturbée, aux non-descripteurs attachés et éventuels termes spécifiques.

Enfin, mettre en place les renvois adéquats a été très important et pour cela il a fallu étudier le vocabulaire employé par les utilisateurs lors de la formulation des demandes. Il faut aider le plus possible les utilisateurs. Si des descripteurs avaient un sens trop proche, ou pas assez précis il a fallu introduire des notes de définition et d'application bien claires, et nous tenons à souligner ce dernier point.

### **Qualité du thésaurus et qualité de l'indexation**

Eurovoc connaît sa troisième édition et un thésaurus qui s'actualise et améliore son niveau de qualité. On sait que ce qui fait la qualité d'un thésaurus est la véracité scientifique de ses concepts traduits en descripteurs et que le suivi des normes de bibliothéconomie est toujours une garantie d'objectivité et d'impartialité.

Toutefois, nous devons considérer un thésaurus comme un discours scientifique et comme tout discours scientifique il est idéologiquement engagé. Un thésaurus est une projection de l'idéologie de ses auteurs, reflet de la réalité sociologique dans laquelle ils s'intègrent. Un thésaurus a la marque du moment historique dans lequel il a été conçu.

Sont demandés aussi des atouts aux indexeurs. Bien indexer est une tâche difficile. L'indexeur doit suivre scrupuleusement les règles de l'objectivité, sélectivité, exhaustivité, spécificité, pertinence et cohérence pour que l'indexation réussisse. L'indexeur doit savoir bien parcourir le thésaurus. Et avant de commencer l'indexation de n'importe quel fond documentaire il doit *sentir la base de données* avec laquelle il travaillera, saisir son profil par des exercices préalables de recherche et seulement après commencer à indexer. Car pour faire une indexation consistante il faut bien s'intégrer dans la politique d'indexation de l'institution et par la suite tester périodiquement son indexation.

Peut-on aboutir à l'indexation parfaite? Le problème de la qualité de l'indexation du côté indexeur est qu'il n'existe pas une seule indexation mais plusieurs indexations. Il faut indexer selon les besoins des utilisateurs, réagir comme ils réagiraient car on n'indexe pas pour soi mais pour eux. Deux indexations différentes d'un même ouvrage peuvent se justifier selon les différentes optiques du centre utilisateur et les spécificités du fond documentaire.

D'ailleurs construire un thésaurus, discuter de sa structure, améliorer, actualiser, mettre de multiples expériences en commun est une *thérapie de groupe* très valable, soit au niveau personnel, soit au niveau de l'institution elle-même. Une entreprise/institution doit être un moyen de communication. Au niveau des institutions, il faudrait instaurer une politique d'indexation par le dialogue, avec partage d'expériences et d'opinions. Les retrouvailles des indexeurs Eurovoc, l'échange d'expériences, sont un moyen de bien indexer et de se sentir bien dans sa peau.

L'effort de qualité n'est pas l'effort d'une seule personne. La qualité (des outils, méthodes et techniques) doit être potentialisée au niveau d'un groupe, d'une institution. Il est conseillé un audit régulier de la base de données pour voir si le thésaurus satisfait aux exigences demandées relatives à la qualité. Et la qualité est surtout un bon service rendu au demandeur d'information, réponses pertinentes avec un bas coefficient de *bruit* et de *silence* dans toutes les recherches car c'est à la recherche qu'on peut évaluer la qualité d'un thésaurus et la cohérence d'une indexation.

BIBLIOGRAPHIE

COMMUNAUTÉS EUROPÉENNES

*Thésaurus Eurovoc*. 3<sup>e</sup> édition. Luxembourg: Office des publications officielles des Communautés européennes, 1996. 3 vol.

DÉGEZ-VATAIRE, D.

*Séminaire pratique d'indexation*. Luxembourg: Parlement européen, 5-7 octobre 1994.