Evaluation de dispositifs d'accès aux ressources terminologiques multilingues électroniques

Mabrouka El Hachani, Fidelia Ibekwe-Sanjuan

Evaluation de dispositifs d'accès aux ressources terminologiques multilingues électroniques

Mabrouka EL HACHANI Fidelia IBEKWE-SANJUAN
ELICO EA 4147
Université Jean Moulin, Lyon 3
6, cours Albert Thomas
69355 Lyon, Cedex 08.
el-hachani@univ-lyon3.fr ; ibekwe@univ-lyon3.fr

Résumé
De nombreuses ressources terminologiques mono et multilingues sont mises à la disposition des usagers et en particulier des professionnels de l'information. Ces ressources, matérialisées aujourd'hui sous forme de dispositifs techniques électroniques servent d'outils de médiation entre les usagers et l'information à traiter. Ils deviennent de fait un dispositif d'appropriation des savoirs.
Nous examinons la structure de ces dispositifs en ce qui concerne les modalités d'accès (options de recherche), les caractéristiques terminologiques qu'ils proposent pour un terme, les parcours d'exploration de l'information qu'ils autorisent. L'objectif à terme est de voir comment ces dispositifs facilitent ou non l'appropriation des connaissances dans un contexte multilingue et plus précisément, comment ils facilitent la recherche des équivalences inter-langues dans le cadre d'exercice d'une activité professionnelle. Cette première étude fait ressortir un certain niveau d'hybridation dans les ressources, à savoir qu'il n'est plus aisé que qualifier telle ressource d'un dictionnaire électronique, telle autre d'une base terminologique, d'un thésaurus ou d'ontologie parce qu'une même ressource, conçue entièrement sous support électronique ou lors de sa migration vers une forme électronique, va mêler des structures et fonctionnalités venant de plusieurs types de ressources. Le modèle hybride, celui d'un dispositif intégré apparaît comme la piste privilégiée pour répondre aux besoins des différents professionnels travaillant dans un contexte multilingue.

Mots-clés : appropriation des savoirs, ressource terminologique électronique, multilinguisme.

Abstract
Many multilingual terminological resources are available today for information professionals and for the general public. Implemented as electronic devices, these resources serve as a mediation tool between users and the information they need, thus making them knowledge acquisition devices. We analyse the structure of some resources, the type of information they offer for a term, the search options as well as the results exploration modalities. The aim is later to assess how these devices contribute towards knowledge acquisition and more precisely, how the terminological resources enable information professionals find equivalent terms in a multilingual framework. The current survey brings to light a certain level of hybridisation among resources making it more difficult to qualify a resource using the usual and classical frontiers (dictionary, thesaurus, ontology, terminology base). This is because, some resources, in the course of their design as electronic devices or during their migration to an electronic form adopted forms and structures issuing from different types of resources. The hybridisation of terminological resources thus appears as the favoured avenue to meet the needs of today's information professionals for knowledge acquisition in a multilingual framework.

Keywords: knowledge acquisition, electronic terminological resource, multilingualism.

1. Introduction

De nombreuses ressources terminologiques électroniques multilingues sont mises à la disposition des usagers et en particulier des professionnels de l'information. Ces ressources, matérialisées sous forme de dispositifs techniques électroniques, servent d'outils de médiation entre les usagers et l'information à traiter. Ils

deviennent de fait un dispositif d'appropriation des savoirs.
Dans sa thèse, (El Hachani, 2005) avait constaté que les normes de construction de thesaurus multilingue[1] ne donnent pas véritablement de solutions pour résoudre tous les problèmes conceptuels posés par la recherche des équivalences multilingues. (El Hachani, 2005) avait également montré que les indexeurs s'appuyaient, non seulement sur leurs propres connaissances pour rechercher des équivalences interlangues, mais également sur des connaissances externes issues de ressources multilingues hétérogènes (dictionnaire, lexique, banque terminologique, encyclopédie).
Nous examinons la structure de ces dispositifs en ce qui concerne les modalités d'accès (options de recherche), les caractéristiques terminologiques qu'ils proposent pour un terme, les parcours d'exploration de l'information qu'ils autorisent. L'objectif, à terme est de voir comment ces dispositifs facilitent ou non l'appropriation des connaissances dans un contexte multilingue et plus précisément, comment ils facilitent la recherche des équivalences interlangues dans le cadre d'exercice d'une activité professionnelle. Des études précédentes (Bertrand 1993 ; Kolmayer 1997 ; Denecker 2002 ; El Hachani 2005) ont souligné la dimension cognitive dans cette relation de médiation entre les ressources électroniques et l'usager. Ces études ont permis de comprendre le processus cognitif des usagers devant des dispositifs techniques et informationnels.
Le reste de ce papier est organisé comme suit : la section §2 synthétise les informations conceptuelles proposées par chaque ressource étudiée ainsi que les modalités de recherche et parcours proposés ; la section §3 dégage des observations à partir des analyses effectuées précédemment et la section §4 trace les perspectives d'avenir.

2. Ressources terminologiques multilingues : état des lieux

Nous avons étudié la structure de plusieurs ressources multilingues en prenant en compte les critères suivants :

1. leur disponibilité sous forme électronique : ces ressources doivent pouvoir être accessibles et interrogeables depuis une interface Web,
2. étude d'un panel représentatif de ressources : nous avons essayé, dans la mesure du possible, d'étudier un exemple significatif d'un même type de ressource (dictionnaire, thésaurus, base terminologique, ontologie, base sémantique lexicale,…),
3. le multilinguisme : les ressources étudiées doivent couvrir plus de deux langues.

Cet ensemble de critères nous a permis de sélectionner cinq ressources qui sont les suivantes : la base ontologique GENOMA[2], le portail terminologique TermSciences[3], le dictionnaire et base sémantique lexicale multilingue Alexandria[4], les thesaurus multilingues Eurovoc[5] et UNBIS[6]. Notre premier travail a été de faire un « inventaire » des informations que nous proposent ces ressources et la manière dont elles sont structurées (les types de relations entre les entités). Pour cela, nous avons élaboré une grille d'analyse des informations et fonctionnalités proposées par les différentes ressources. L'élaboration de cette grille a suivi une approche descendante (*bottom-up*), à savoir que ce sont les ressources elles-mêmes qui ont fait émerger la grille en fonction des fonctionnalités dont elles disposaient. Ceci nous a permis de classer les types d'informations disponibles en cinq catégories :

1. informations générales (type de la ressource, institution d'origine et autres informations pratiques),
2. informations linguistiques (morphologique, lexicales) et relations conceptuelles proposées pour un terme,
3. contexte discursif du terme (fonction de concordancier, extrait de textes le contenant),
4. couverture multilingue et informations sur les équivalences multilingues
5. options de recherche et modalités de navigation (parcours proposés).

[1] Il s'agit des normes :
ISO 2788-1986 - Principes directeurs pour l'établissement et le développement des thésaurus monolingues.
ISO 5964-1985 - Principes directeurs pour l'établissement et le développement des thésaurus multilingues.
[2] http://genoma.iula.upf.edu:8080/genoma/index.jsp
[3] http://www.termsciences.fr/article.php3?id_article=38
[4] http://www.tv5.org/TV5Site/alexandria/index.php
[5] http://europa.eu/eurovoc/
[6] http://unhq-appspub-01.un.org/LIB/DHLUNBISThesaurus.nsf/$$searchf?OpenForm

Chacune de ces catégories d'informations est déclinée en autant de points à renseigner dans la grille que d'informations disponibles dans les cinq ressources étudiées. Nous analysons ci-dessous les résultats de ce premier travail d'état des lieux. Pour des raisons pratiques, nous regroupons les catégories d'informations dans la grille en trois parties : (i) les informations de nature linguistique et conceptuelle fournies pour un terme, (ii) les informations multilingues disponibles pour un terme, (iii) les options de recherche et de navigation au sein des résultats. Nous laisserons de côté les informations générales, à caractère pratique qui ne nécessitent aucune analyse.

2.1 Informations linguistique et conceptuelle disponibles pour un terme

Nous allons synthétiser les informations présentes dans ces ressources selon trois axes : la qualification de la ressource, la différenciation terme / concept, les informations linguistiques et les relations conceptuelles proposées pour un terme.

2.1.1. Qualification de la ressource

Les ressources étudiées se présentent chacune sous un type donné même si comme on le verra plus tard, les informations et fonctionnalités qu'elles proposent ne permettent plus de tracer des frontières nettes entre elles. Genoma est présentée comme une « *banque termino-ontologique* » sur le génome humain. TermSciences est présenté comme un « *portail terminologique multidisciplinaire* » contenant les termes contrôlés issus de nombreuses autres ressources dont les deux lexiques (PASCAL, FRANCIS) de l'INIST, les termes issus du thésaurus MESH, un dictionnaire de l'INRA sur les biotechnologies et le thesaurus de la banque de données en santé publique (BDSP). Le terme même de portail change inévitablement le mode de représentation que l'on peut avoir de l'outil ainsi que les attentes en terme d'usage que nous pouvons également en avoir. UNBIS et Eurovoc sont des thesauri multilingues. Eurovoc est un thesaurus multidisciplinaire et couvre tous les domaines intéressant les activités des institutions européennes. UNBIS est un thesaurus multidisciplinaire utilisé par les différents organismes des Nations Unies pour indexer leurs documents. Alexandria est la ressource la plus difficile à qualifier. Dispositif protéiforme, ses auteurs la présentent tour à tour comme un appartenant à plusieurs famille dont les logiciels d'aide contextuelle, les logiciels d'aide à la compréhension [7]» et également comme un « agent intelligent [8]». Alexandria est à mi-chemin entre un dictionnaire multilingue et une base sémantique multilingue, *à la WordNet*. En effet, elle a été réalisée en s'appuyant sur la structure de WordNet pour la partie anglaise et sur le Dictionnaire Intégral de la société Mémodata[9] pour la partie française.

De ce premier tour d'horizon sur les qualifications, il apparaît que nous avons dans notre panel deux thesaurus multilingues assez classiques et trois ressources termino-ontologiques multilingues. Nous avons cherché ensuite à cartographier les types de relations linguistiques et conceptuelles que ces ressources proposaient entre une entrée terminologique et les autres. Une première question qui s'est posée à la lumière de certaines informations dans ces ressources est celle de la différentiation entre une entrée terminologique de la ressource (terme) et le concept désigné.

2.1.2 Différenciation entre terme et concept

Visiblement, les ressources se différencient en deux catégories sur ce point : celles qui introduisent une nuance entre « terme » et « concept » (Genoma, TermSciences) et les thesaurus (UNBIS, Eurovoc) ou lexicologique (Alexandria) qui ne le font pas. Les ressources qui se réclament plus explicitement de la tradition ontologique (ingénierie des connaissances) distinguent entre le terme (la matérialisation linguistique du concept) et le concept (l'idée ou l'objet du monde réel, le référent du terme). Cette distinction a pu se concrétiser au niveau informatique (dans le modèle de bases de données utilisé pour bâtir la ressource) mais elle demeure peu perceptible au niveau de l'utilisation de ces ressources. C'est notamment le cas pour la

[7] http://www.tv5.org/TV5Site/alexandria/entretien.php.
[8] http://www.tv5.org/TV5Site/alexandria/index.php
[9] Présentation de la ressource à http://www.memodata.com/2004/fr/dictionnaire_en_ligne/index.shtml

banque termino-ontologique Genoma, où les informations terminologiques données pour un terme distinguent entre le terme en catalan et le « concept » qui n'est que la traduction du terme en anglais (voir figure 1). Cette distinction peut donc apparaître artificielle pour l'usager. En effet, l'identifiant du concept et sa matérialisation linguistique (le terme) ont souvent la même apparence (même terme). Dans TermSciences, « concept » n'est employé que pour expliquer la structure de la base terminologique mais ce terme n'apparaît pas au niveau de l'exploitation de la ressource. La distinction entre le terme et le concept sert donc à relier l'entrée terminologique au concept qu'elle identifie de manière unique dans la base de données quelle que soit les langues de traduction du terme. Si elle est totalement justifiée au niveau théorique linguistique, la distinction entre le terme et le concept dans ces deux ressources apparaît néanmoins comme une précision déroutante pour le non-initié car il ne perçoit finalement pas la différence qui est faite dans les résultats des recherches et dans les parcours proposés.

Figure 1. Genoma : informations terminologiques en réponse à une recherche sur la chaîne « gene ».

2.1.3 Informations linguistiques et relations conceptuelles disponibles dans les ressources

Seules les ressources de type termino-ontologiques (Genoma, TermSciences) et lexicographiques (Alexandria) proposent une définition pour chaque terme. Cette information est génralement absente dans les thesauri et, Eurovoc et UNBIS ne dérogent pas à la règle. De même, ces trois premières ressources se différencient des autres par la présence des informations linguistiques : la catégorie morphologique et le genre d'un mot, les informations flexionnelles et derivationnalles avec des étiquettes telles que [Nominalisation, dérivé] dans Alexandria pour marquer la relation lexicale entre «*juger*» et «*jugement* ». Genoma est la seule ressource à vraiment proposer la fonction de concordancier. Etant donné que cette ressource est couplée à un corpus (qui a permis d'extraire et de modéliser les termes du domaine dans une ontologie), il était plus aisé d'afficher les extraits de textes dans lesquels apparaît un terme. Alexandria ne propose que des exemples de phrase pour illustrer tel ou tel sens d'un terme, tel un dictionnaire de langue.

Toutes les ressources proposent des relations conceptuelles générales d'hyperonymie (TG), d'hyponymie (TS) et de synonymie. Cependant, seuls les thesaurus précisent le statut préférentiel d'un terme avec la mention « EM/EP ». Seule Genoma explicite la relation de co-hyponymie (celle qui existe entre tous les hyponymes (TS) d'un même hyperonyme (TG)). Genoma est également la seule ressource à proposer des variantes d'un même terme qui ne sont ni ses synonymes, ni ses termes associés (TA) mais des termes qui s'en approchent par des phénomènes de variations linguistiques telles que celles décrites dans Ibekwe-SanJuan (2006). Notons également que les deux ressources Genoma et Alexandria proposent d'autres relations conceptuelles. A titre d'exemple, Génoma propose les relations « *is_process_of, stage_of, occurs_after* » qui sont des spécifiques au domaine de la génomique. Alexandria propose des d'autres

relations conceptuelles qui sont indépendantes d'un domaine mais reflètent un degré de modélisation sémantique plus poussée. Ainsi, « *PersonneQuiFait* » qui est une relation d'agent, relie « *arrêtiste* » et la classe de concepts « *arrêt, jugement, sentence, verdict* ».
Concernant le nombre de relations conceptuelles proposées, les thesaurus apparaissent comme plus synthétiques que les ressources termino-ontologiques. En effet, les premières condensent sous un même nom, des relations conceptuelles différentes mais au prix d'un certain taux d'ambiguïté voire d'incohérence dans les relations entre les termes. Par exemple, dans le thesaurus Eurovoc, sous la relation d'équivalence (EM/EP) sont rangées plusieurs types de relations dont «*la synonymie vraie, la quasi-synonymie ou voisinage de sens, l'antonymie, ou opposition de sens ou l'inclusion*[10]». La relation associative recouvre également une multitude de relations telles que la causalité, l'instrumentation, la hiérarchie (en cas de poly-hiérarchie), la concomitance ou encore les matériaux utilisés. Il est vrai que la relation d'association sert traditionnellement de relation « fourre-tout ».

2.1.4 Cas d'Alexandria

Une place à part doit être faite à Alexandria concernant les relations conceptuelles proposées. Alexandria se différencie assez des autres ressources par la grande variété de relations qu'elle propose et par un effort plus poussé de modélisation des relations entre objets du monde réel. Alexandria est davantage à une base sémantique lexicale (réseau sémantique) structurant les concepts du domaine général dans une langue et établissant les équivalences dans une deuxième langue. Selon les informations que nous donne le site de Mémodata[11], Alexandria, qualifié de dictionnaire électronique, recouvre dans son contenu à la fois un dictionnaire de définitions, un dictionnaire de synonymes, un dictionnaire de traductions vers 22 langues, des entrées analogique et onomasiologique («*dictionnaire allant des idées vers les mots»*), des dérivés sémantiques, la navigation dans un réseau lexical à la WordNet et des éléments d'ontologie. Les relations macro-hiérarchiques telles que « Domaine, Thème, ClasseHyper » côtoient des relations hiérarchiques au niveau d'un terme : [Hyper →hyperonymie], [Spéc. →spécifique].
Désirant vérifier la structure de représentation des connaissances sous-jacentes à cette ressource, nous avons pris contact avec son auteur[12] pour obtenir plus de détails. Alexandria, structurée au départ comme un treillis (appelé le Dicologique) a évolué progressivement ver un hypergraphe (aujourd'hui le Dictionnaire Intégral).

Un des principes fondamentaux du modèle sémantique adopté est d'autoriser l'héritage multiple (poly-hiérarchie) d'un objet de plusieurs classes, d'où la structure en treillis. La raison ayant motivé ce choix est le constat qu'un même objet ou concept peut avoir plusieurs hyperonymes (TG) en fonction de ses différents sens ou points de vue. Il peut au contraire ne pas avoir d'hyperonyme du tout. L'idée est alors de construire des familles d'objets appartenant à une même classe (*synsets*), un même objet pouvant appartenir à plus d'une classe. Pour illustrer la structure d'organisation des connaissances adoptée dans Alexandria, considérons cet exemple que nous donne son auteur : terme ou l'objet « *yen* » appartient à la classe « *monnaie du Japon* » qui est donc son hyperonyme, désigné par la relation [Classe]. Cette classe est à son tour incluse dans une classe plus générale de « *Monnaie d'Asie* » qui en est donc l'hyperclasse, désignée par la relation [ClasseHyper]. Cette dernière est incluse dans l'hyperclasse «*Monnaie* » et ainsi de suite. Compte tenu de leur caractère très générique, les hyperclasses ne précisent pas la nature exacte de la relation entre les objets qu'elles contiennent et apparaissent plus comme une relation paradigmatique. Les concepts de type [ClasseHyper] peuvent contenir trois types d'informations :
d'autres hyperclasses (exemple « *monnaie d'Asie* », une hyperclasse contenue dans « *Monnaie* »)
les mots désignant les concepts génériques (exemple « *monnaie* »)
des termes spécifiques, instances d'une même classe d'objets (exemple *dollars américain, yen, livre sterling, euros*,…).

[10] *Voir* http://europa.eu/eurovoc/sg/sga_doc/eurovoc_dif!SERVEUR/menu!prod!MENU?langue=FR.
[11] *Voir* http://www.memodata.com/2004/fr/dictionnaire_en_ligne/index.shtml.
[12] . Nous rendons compte ici d'un échange de courriels avec Dominique Dutoit, chercheur au CNRS.

Pour illustrer la mise en œuvre de la poly-hiérarchie dans Alexandria, considérons maintenant la relation qui unit « *yen* » à l'objet « *Japon* » qui, par rapport aux termes déjà énumérés ci-dessus est une entité nommée (noms de personnes, de lieux, de temps). Pour pouvoir marquer cette autre dimension de relation entre « *yen* » et « *Japon* », les auteurs d'Alexandria ont décidé de créer une autre relation appelée [Thème] qui désigne des classes de conteneurs. La relation de « thème » apparaît, après la relation « ClasseHyper », comme une deuxième relation de type paradigmatique et permet de visualiser la relation entre objets d'un point de vue différent. Selon cette même logique, l'hyperclasse « *monnaie* » constitue également un « Thème » du même nom.
Une autre relation désignée par [DomaineCollocation] regroupent des classes d'objets d'une même nature grammaticale mais qui ne sont pas dans une relation définitoire avec l'objet considéré (ne permettent pas de le définir). Dans l'exemple précédent, les concepts de « *monnaie* » et le thème de « *Japon* » sont définitoires pour « *yen* » alors que les concepts de « *prix* » et de « *température* » ne le sont pas pour l'objet « *monter* ».
Les entrées (termes) d'Alexandria peuvent être des mots simples (*dispositif*) ou bien des pluritermes (*procédure judiciaire, arrêt de mort*). Cette ressource n'est donc pas assimilable à un dictionnaire qui n'aurait que des entrées atomiques, constituées des mots d'une langue. L'utilisateur rentre par une recherche simple sur un terme. Alexandria affiche les différents sens connus du terme ainsi qu'une liste de pluritermes le contenant (relation d'inclusion lexicale). Une fois dans l'environnement lexicographique du terme, ces différents sens peuvent donner lieu à des classes de sens composés de plusieurs mots (pluritermes) voire à des expressions qui ne sont pas des termes. Ainsi, parmi les sens répertoriés en français pour le terme « *dispositif* », on trouve un sens médical « *moyen pour parvenir à un résultat* » (voir figure 2) qui relève par ailleurs d'une classe de termes [Classe]. Etant donné que toutes les entrées sous un terme sont cliquables, un clic amène l'utilisateur dans l'environnement sémantique de cette classe qui se trouve en réalité être une sous-classe de sens parmi le thème plus large dénommé « *moyen pour parvenir à un résultat déterminé* » (voir figure 3).

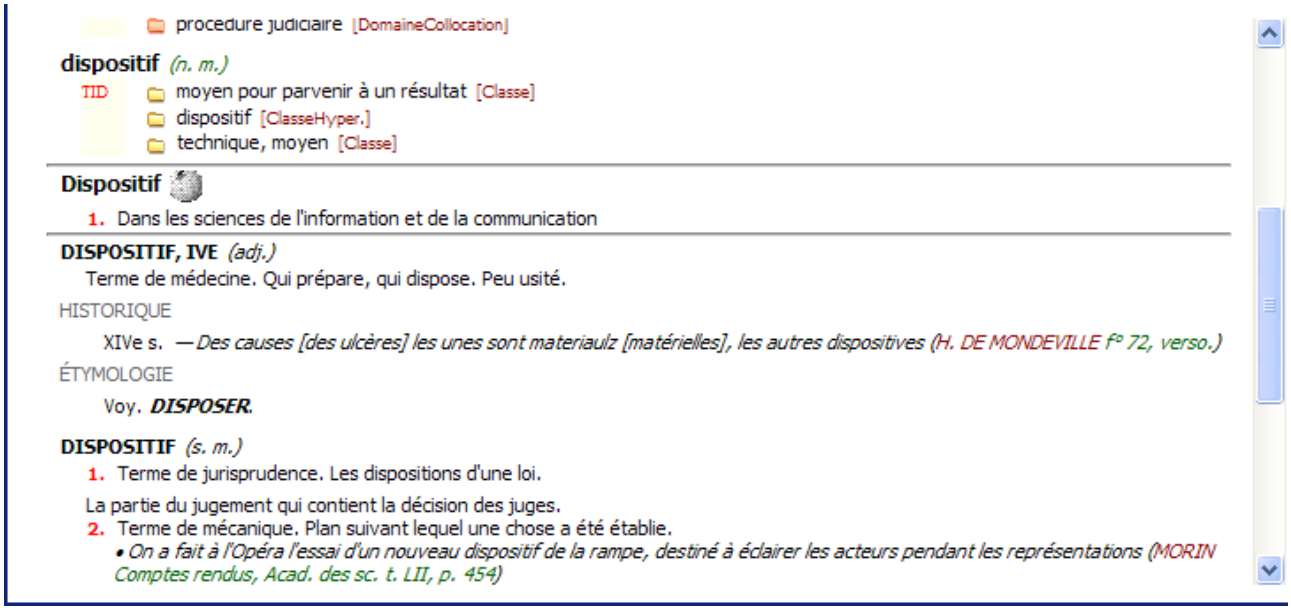

Figure 2. Alexandria : environnement sémantique du mot «*dispositif* ».

Figure 3. Alexandria : environnement sémantique de la classe «*moyen pour parvenir à un résultat*».

La structure arborescente sous une entrée dans Alexandria est rendue perceptible par l'usage des retraits et un usage fort astucieux des couleurs. Notons enfin une singularité d'Alexandria au niveau de la présentation des classes d'équivalences (*synsets*). Une fois dans l'environnement sémantique d'un terme, cette interface propose à l'utilisateur d'accéder à des classes de concepts proches. Ces classes de concepts sont matérialisées comme des listes de termes solidarisées d'une même classe (*synsets*), indiquée par une pastille rouge (voir figure 3 ci-dessus la liste commençant par « *pratique* »). Un clic sur une liste affiche l'environnement sémantique de la classe avec ses propres liens conceptuelles (classe d'appartenance ou de dépendance), relations plus spécifiques au terme (hyper, spéc.) si elles existent.

2.2. Informations sur les équivalences multilingues d'un terme

Nous avons étudié la manière dont la ressource propose la recherche d'une équivalence dans une autre langue, le type d'informations fournies ainsi que la possibilité du choix de la langue pour l'interface. Les niveaux d'informations affichées dans les langues cibles sont variables. Cela permet de distinguer entre deux catégories de ressources : thesaurus et bases termino-ontologiques.

2. 2.1 Informations affichées par les thesaurus

Schématiquement, les thésaurus proposent les relations conceptuelles génériques usuelles dans les autres langues couvertes : l'équivalence du terme, ses génériques et spécifiques (TG/TS), la relation de synonymie (EM/EP), les termes associés (TA) ou encore une note d'usage (ou d'application). Ceci est notamment vrai pour UNBIS. Ce thesaurus propose les équivalences dans les six langues officielles des Nations-Unies (anglais, arabe, chinois, espagnol, français et russe). L'utilisateur peut choisir la langue source soit par une recherche soit par exploration des listes thématique (micro-thesaurus) ou alphabétique mais il ne peut pas choisir de visualiser une langue cible parmi les six officielles. Les informations sur les équivalences dans les six langues sont affichées en même temps sur un même écran (voir figure 4 ci-dessous).

Dans le thesaurus *Eurovoc*, la recherche d'un terme affiche l'environnement sémantique du terme (ses TG, TS et termes associés) se fait dans une langue source. Les menus de l'interface bascule alors dans la langue source choisie. Les langues cibles ont toutes le même statut, à savoir qu'à chaque descripteur dans une langue correspond obligatoirement un descripteur dans chacune des autres langues. Il y a une relation bi univoque entre descripteurs de différentes langues mais il n'y a pas d'équivalence interlangue pour les non-descripteurs compte tenu du fait que les langues sont caractérisées par une richesse lexicologique qui varie d'un domaine à l'autre. Bien que l'étendue linguistique de cette ressource couvre un nombre très élevé de langues (23), curieusement, la recherche d'équivalences interlangues n'est possible que par une fonction « liste multilingue » qui permet de sélectionner entre 2 et 4 langues à afficher (voir figure 5 ci-après). L'interface génère ensuite une liste de ses descripteurs dans ces 4 langues mais par ordre alphabétique de l'ensemble des descripteurs du thésaurus. Il ne semble pas y avoir la possibilité pour l'utilisateur de chercher un terme dans une langue cible et d'obtenir ses équivalences dans quelques unes ou dans les 22 autres

langues. Par ailleurs, il ne semble pas possible de fixer l'étendue de la liste générée. Les possibilités de recherche d'équivalences multilingues semblent très pauvres dans cette ressource au vu des nombreuses possibilités qu'offre la mise sous forme électronique de la ressource.

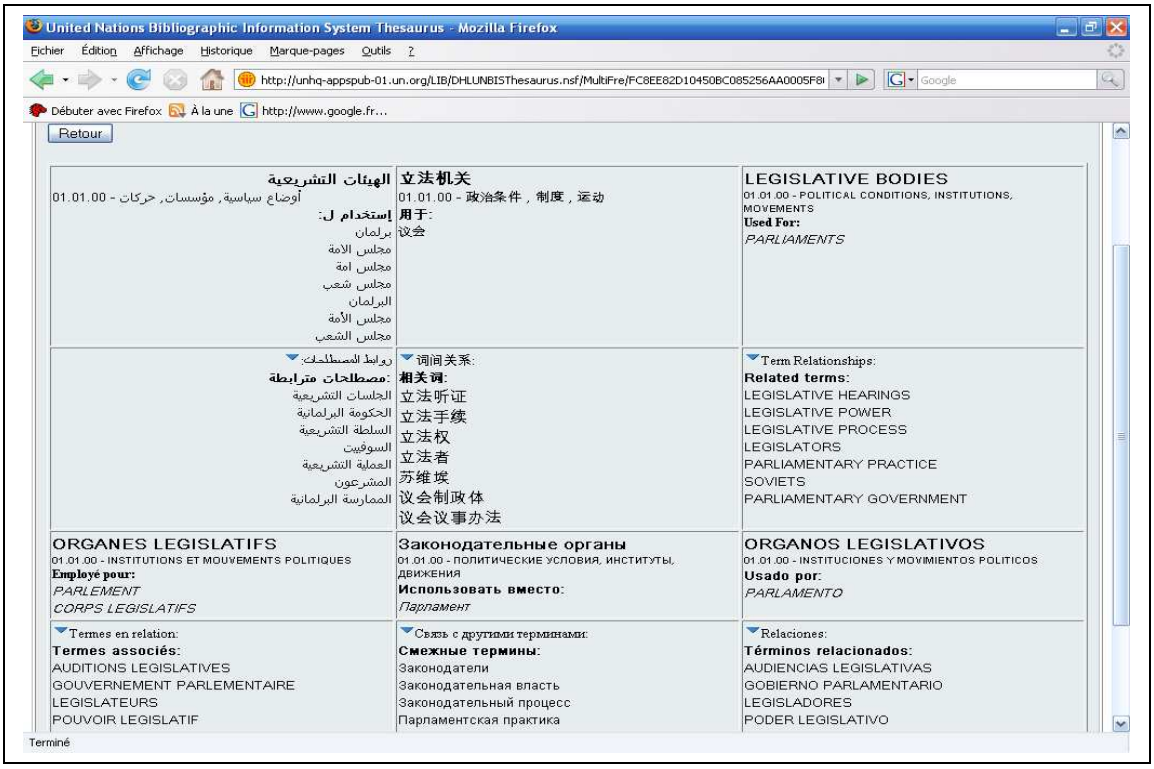

Figure 4. Thesaurus UNBIS. Environnement multilingue d'un terme.

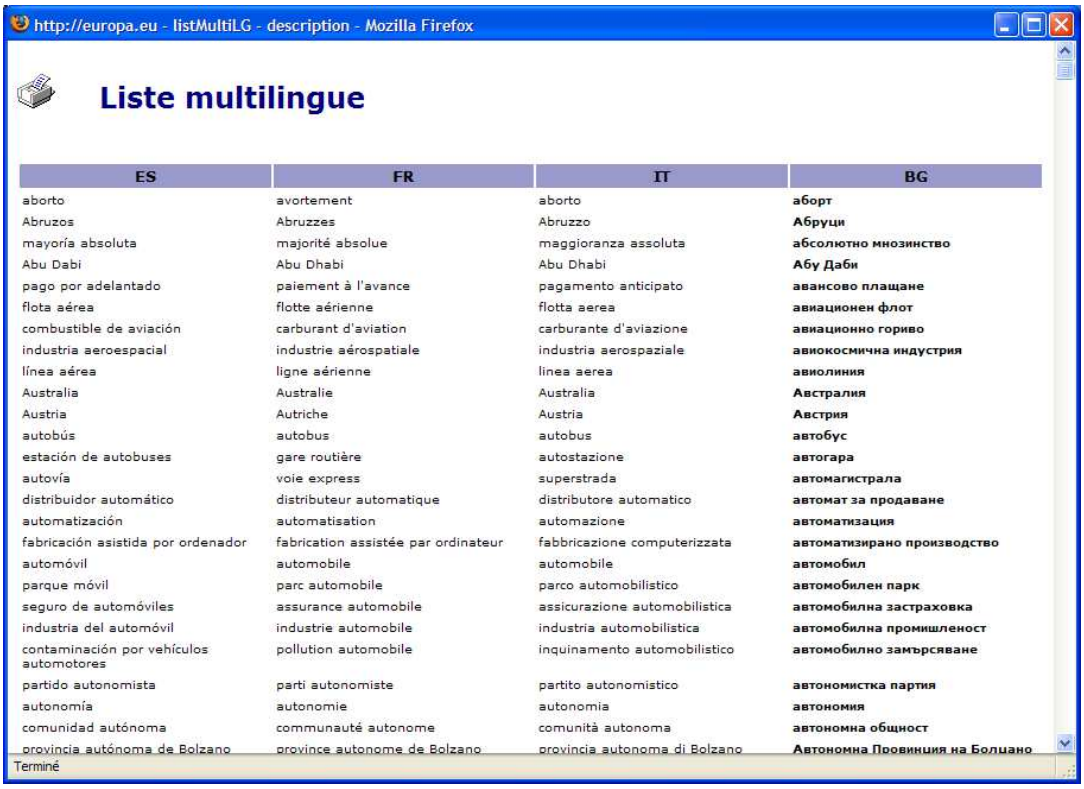

Figure 5: Eurovoc : liste alphabétique des descripteurs en quatre langues.

Une fois la recherche effectuée, l'équivalence interlangue peut être visible par le lien « traduction » dans la barre de menu, et là apparaît une liste des équivalents dans les 23 langues identifiées par des icônes. Il est

cependant à remarquer que seul le descripteur en tête de la hiérarchie est traduit. L'utilisateur à la possibilité d'afficher une autre version linguistique de cette arborescence en visualisant non seulement la branche du thesaurus dans la langue cible mais également d'afficher toute l'interface dans la langue ainsi sélectionnée.

2. 2.2 Ressources de termino-ontologiques

Elles offrent globalement plus de possibilités aussi bien au niveau des options de recherche, celui des informations multilingues proposées que de parcours dans la ressource. La recherche d'équivalence interlangue dans Alexandria passe par une fonction traduction qui permet de sélectionner la langue cible d'un terme mais on ne peut visualiser que des équivalences bilingues. Donc même si cette ressource couvre 22 langues, on ne peut travailler que sur deux langues à la fois. L'équivalence dans la langue cible donne les différents sens du mot mais les autres informations sémantiques sont données dans la langue source. A titre d'exemple, la recherche de l'équivalence du terme « *avocat* » en anglais fournit les différents sens répertoriés de ce terme : {*avocat : barrister, lawyer, solliciter ; avocat (n) : advocate, advocator, avocado (biologie),…*}. Tous les termes ici sont cliquables et permettent de basculer dans l'environnement sémantique du terme dans la langue cible. Les autres informations conceptuelles concernent le terme sont fournies dans la langue source. Ainsi, nous avons dans ce cas précis les relations d'inclusion lexicale (termes contenant le mot *avocat* : *avocat aux conseils, avocat commis, sandwich à l'avocat…*). Cependant lors de l'exploration, nous avons constaté qu'il n'y avait pas la même couverture pour toutes les langues et pour toutes les entrées du dictionnaire. Cela signifie également qu'il n'y a pas la même couverture pour chaque paire de langues, ainsi les informations ne seront pas complètes pour toutes, comme pour l'exemple de recherche « sciences sociales » ou nous n'avons pas d'équivalent en espagnol, en arabe et dans d'autres langues, en revanche, nous l'avons trouvé pour l'italien et le portugais.

La base Genoma étant une ressource spécialisée, n'a pas la même ambition de couverture multilingue que les autres ressources. Cela est certainement du à son contexte d'élaboration. Elle propose une recherche uniquement vers trois langues (espagnol, catalan et l'anglais), de plus, il est possible de choisir la langue de l'interface utilisateur. En revanche, la recherche permet de choisir la langue source celle de la requête, les informations terminologiques affichées par la suite seront dans cette langue-là. L'équivalence interlangue est disponible à partir de l'icône correspondant. Outre l'équivalence interlangue, d'autres informations terminologiques, lexicographiques, contextuelles et documentaires sont disponibles à travers les modules correspondants (base terminologique, base textuelle, base documentaire, base factographique). Nous avons remarqué qu'il était possible de choisir la langue de requête pour les bases terminologique et textuelle et non pour les deux autres qui restent dépendantes de la langue de l'interface.

TermSciences propose le choix entre quatre langues d'interrogation. Ainsi, une requête sur un terme en français dans *TermSciences* donne ses équivalents en anglais, espagnol et allemand mais la définition du terme n'est fournie qu'en anglais. Ensuite, il fonctionne comme *UNBIS*, c'est-à-dire qu'il donne l'équivalence interlangue pour les langues présentes mais contrairement à *UNBIS* qui propose toujours les équivalences interlangues dans les six langues, la quatrième langue, l'allemand apparaît peu. Nous avons ainsi remarqué une grande disparité entre les langues pour cet outil, il semble y avoir une grande différence lorsque l'on change de langue pour un même terme. Ainsi pour le terme « génome » en français, une liste de 73 réponses s'affiche. En reprenant l'équivalent anglais en tant que langue d'interrogation c'est à dire « genome » on note seulement 31 réponses. Nous avons également remarqué que la partie combinaison linguistique change lorsque l'on modifie les combinaisons de langues pour un même concept, en revanche la partie de l'interface correspondant au résultat et donc à la définition de ce concept, ne change pas.

2.3. Options de recherche et parcours proposés pour explorer les résultats

Nous avons abordé indirectement cette question dans les sections consacrées aux relations conceptuelles (§2.1.3) et aux informations multilingues (§2.2). Nous mettrons l'accent ici sur d'autres mécanismes d'affichage mise en œuvre dans ces ressources qui permettent une meilleure appréhension du sens d'un terme dans un environnement multilingue. Toutes les ressources proposent au moins un accès par recherche

directe d'un terme. Cette recherche peut porter sur le terme exact (Alexandria) ou sur une partie du terme (Genoma, TermSciences, UNBIS, Eurovoc), ce qui autorise une recherche approximative moyennant les opérateurs usuels (booléens, troncature, syntaxique, adjacence). Ceci est particulièrement appréciable lorsqu'on désire voir tous les termes contenant une partie d'un autre terme ou si on ignore l'orthographe exacte d'un terme. Seule Genoma propose en plus la recherche d'un terme selon ses attributs linguistiques (catégorie morphologique et lemme) et statistiques (fréquence dans le corpus). TermSciences, UNBIS et Eurovoc proposent de rechercher un terme selon son statut préférentiel, marquant ainsi leur héritage de la tradition documentaire des thésaurus. UNBIS et Eurovoc proposent par ailleurs une recherche par micro-thésaurus. Ceci permet à un utilisateur de rentrer directement dans l'arborescence du thésaurus par un domaine (micro-thésaurus) et de visualiser la structure arborescente (relations de TG, TS) et horizontale (relation de TA). TermSciences et UNBIS proposent également un accès par liste alphabétique.
TermSciences offre la possibilité d'étendre la recherche sur des moteurs de recherche grand public (Google). TermSciences et Alexandria proposent de consulter la source d'un terme lorsque l'entrée provient d'une autre ressource (MeSh, CISMEF, INRA, PASCAL, FRANCIS, Dictionnaire Intégral, WordNet). De plus, *Alexandria* permet d'étendre la recherche d'information sur d'autres ressources en ligne de type enclyclopédique signalées par un logo (Wikipédia, Littré). Il semble également possible d'utiliser le portail TermSciences comme outil d'indexation vers le web2 en référençant les concepts trouvés vers des sites tels que *digg* ou *deli.icio.us* (rubrique « *Référencer ce concept sur …* » avec icônes correspondants aux sites), seulement pour pouvoir entrer ce concept, il faut se créer un compte chez un de ces prestataires.

Concernant les options d'affichage et les parcours proposés, la plupart des ressources adoptent la représentation sous forme d'arbre permettant ainsi d'apercevoir les relations hiérarchiques (TG, TS) autour d'un terme. Cette présentation est accompagnée dans TermSciences et dans Genoma par des informations sur les définitions et les équivalences multilingues. Seule Genoma propose la fonction de concordancier (KWIC) pour l'affichage des contextes d'emploi d'un terme. Dans l'interface de UNBIS, les résultats d'une recherche sur un terme affiche un tableau donnant les termes environnants à celui saisi (jusqu'à 500). A ce premier niveau d'affichage, l'interface affiche les non descripteurs (équivalents non autorisés). En cliquant sur un terme, l'interface montre son environnement multilingue dans un tableau synthétique (figure 4 ci-dessus).
Les deux figures ci-après montrent les modes d'affichage de l'environnement multilingue d'un terme dans TermSciences.

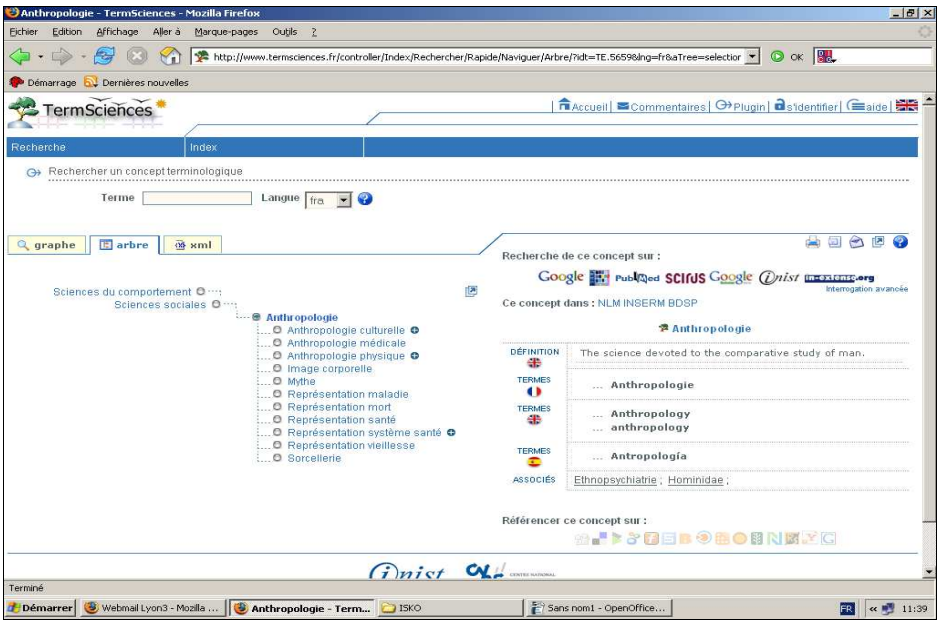

Figure 7: TermSciences : affichage « arbre » et possibilités d'extension de la requête par l'indexation sociale (*digg*, *deli.icio.us*)

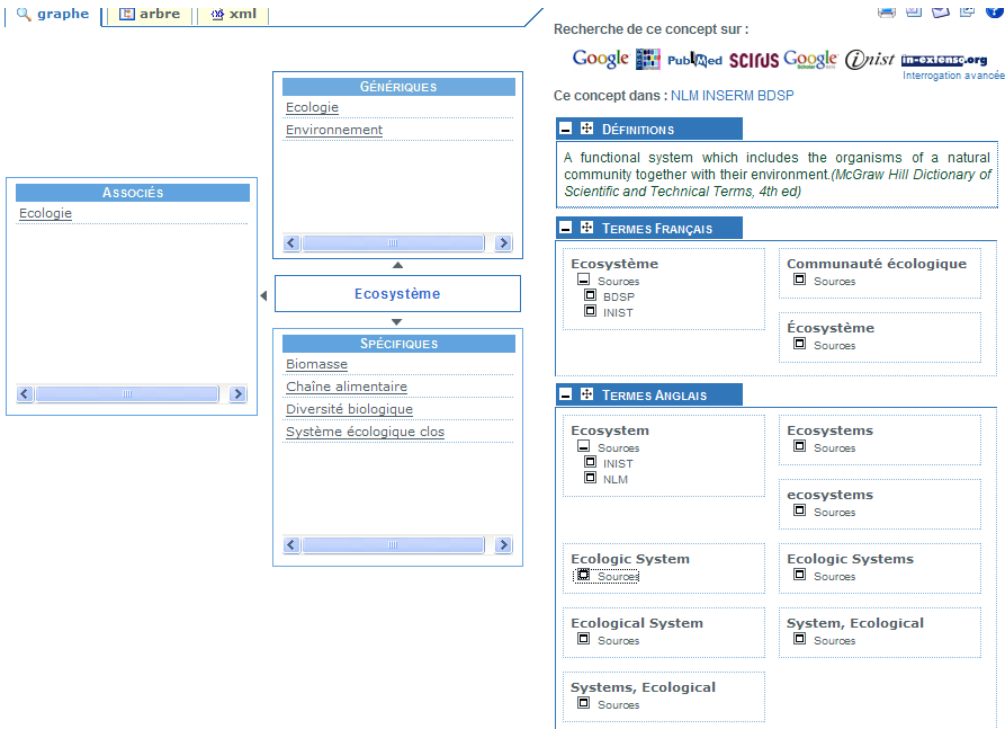

Figure 8. Affichage de type graphe de l'environnement du terme « *Ecosystème* ».

Les équivalents multilingues sont présentés dans une autre partie de la fenêtre sans le contexte hiérarchique mais avec la définition du terme en anglais. Ensuite, un clic sur un terme dans l'arbre permet de développer sa propre arborescence. A l'heure actuelle, seule TermSciences propose une sortie au format XML pour une fiche terminologique.

3. Mutation des ressources terminologiques électroniques

3.1. Vers un modèle de ressources hybride ?

Le premier constat qui émerge de ce premier tour d'horizon est que les ressources terminologiques électroniques évoluent vers l'hybridation. Les anciennes frontières classiques permettant de démarquer les ressources par type (lexique, thésaurus, dictionnaire, ontologie) ne sont plus systématiquement opérantes (voir aussi Ghilchrist 2003 pour une discussion). La plupart des ressources sondées, notamment TermSciences, Alexandria et Genoma, offrent des fonctionnalités hybrides mêlant celles d'un dictionnaire, d'un thésaurus, d'une ontologie et des outils d'exploration automatique de corpus, si bien qu'il faut aujourd'hui inventer de nouveaux qualificatifs pour les décrire. D'ailleurs, les dénominations de ces ressources le montre : « *portail terminologique* » pour TermesSciences, « *base de connaissances termino-ontolgique* » pour Genoma et « *outils d'aide contextuelle, dictionnaire électronique multilingue* » ou encore « *agent intelligent* » pour Alexandria. Ces dénominations, non encore consensuelles, sont les signes du flottement que ressentent leurs auteurs dans le choix d'un qualificatif pouvant caractériser au mieux les attributs de leurs ressources. Cette hybridation apparaît de manière assez claire pour les trois ressources Genoma, Alexandria et TermSciences bien que cette dernière reste encore marquée par l'héritage documentaire des thésaurus qui constituent l'essentiel de ses entrées.

De manière générale, les ressources de type thésaurus (UNBIS et Eurovoc) offrent des possibilités de recherche et de navigation assez pauvres dans leurs versions électroniques. Elles n'ont pas véritablement fait la transition vers des ressources terminologiques protéiforme, exploitant toutes les possibilités qu'offrent le passage au numérique. UNBIS et Eurovoc restent très clairement des thésaurus dont l'élaboration est

corsetée par des normes du métier de la documentation.

Parmi les cinq ressources sondées, Alexandria apparaît comme la plus innovante et la plus difficile à étiqueter. Elle n'est ni un dictionnaire électronique, ni une ontologie, ni une base sémantique lexicale, elle est les trois à la fois. Elle facilite l'appropriation du champ sémantique autour d'un terme à l'intérieur d'une même langue grâce à la variété d'informations et de relations fournies pour chaque terme. L'interface facilite assez cette compréhension même si elle n'offre pas une vue globale de la macro-structure (l'arborescence des domaines, thèmes et classes). Cependant, il n'existe pas encore aujourd'hui un modèle hybride unique de toutes ressources terminologiques électroniques. Elles partagent néanmoins des caractéristiques communes dont le socle semble tourner autour d'un noyau de relations conceptuelles (TG, TS, TA, synonymie), d'une représentation arborescente des parties du réseau sémantique et des fonctions d'interrogation. Les modèles cognitifs sous-jacents à ces outils induisent de nouveaux parcours dans les modes d'accès aux connaissances qu'il nous faudra analyser dans une étude ultérieure.

3.2. Vers des ressources multilingues idéales ?

Des études antérieures ont constaté les insuffisances des ressources terminologiques existantes dans le traitement des équivalences multilingues. Différentes propositions pour combler ces lacunes ont alors été émises. (Lyadri, 1997) faisait état, dans un article traitant de la problématique de la traduction français/arabe, des difficultés rencontrées par les utilisateurs du dictionnaire bilingue pour la recherche d'équivalence malgré la mise à disposition d'information lexicographique et sémantique. (Dancette, 2003) évoquait ailleurs l'insuffisance des dictionnaires bilingues dans le travail de traduction et de recherche d'équivalence et la nécessité de disposer d'autres informations pour rendre pertinent le travail du traducteur.

Dès 1989, (Larivière, 1989) préconisait déjà un produit unifié en terminologie et en documentation en évoquant un outil hybride intitulé « thesaurus terminologique » qui pourrait synthétiser les fonctions du thesaurus et celui du lexique terminologique. Il apparaissait au résultat de ce travail que des ressources de type encyclopédique pourraient également trouver leur place dans ce dispositif, dont la nécessité n'est plus à démontrer en contexte multilingue.

L'arrivée du web sémantique a suscité de nombreux questionnements chez les professionnels de l'information (voir par exemple Gilchrist (2003). Une question récurrente est celle de la différence qui existe entre une taxonomie, un thesaurus et une ontologie ?[13] Avec l'arrivée des outils terminologiques hybrides, cette question n'est pas prête de trouver une réponse précise sinon dans la réflexion d'un méta-modèle intégrant tous ces concepts à la fois.

Cette question est sans doute plus large[14]. En effet, (Dyens, 2004) évoquait l'émergence d'une nouvelle structure de connaissances avec l'ère du web, cette structure émergente serait ni linéaire ni verticale mais neuronale. Ainsi appréhender les dispositifs terminologiques techniques comme les ressources en ligne implique un nouveau mode d'usage et une nouvelle manière d'appréhender l'information électronique, son caractère changeant et interactif. (Origgi, 2004) s'interrogeait sur la relation entre le traitement de l'information et les sciences cognitives à travers l'usage des nouvelles technologies et ainsi l'appropriation du savoir et des connaissances par ces outils de médiation. Il s'agit donc bien d'un processus cognitif et de la manière dont ces outils offrent aux usagers une liberté de construction du sens. Il faut pour cela avoir accès à différents types d'information à travers des outils hybrides qui en permettront l'exploration.

4. Perspectives

Ce premier travail d'état des lieux des ressources terminologiques multilingues à l'ère électronique est loin d'être achevé. Il n'a pas été facile de trouver des critères d'analyse communs (la grille). Les dispositifs électroniques étudiés sont des héritières de trois traditions anciennes issues de disciplines scientifiques distinctes : lexicographie-linguistique (dictionnaires), documentation-bibliothèque (thesaurus), ingénierie des

[13] En témoigne la journée d'étude de l'ADBS « *Le web sémantique : de nouveaux enjeux documentaires*, octobre 2003, disponible : http://adbs.fr/uploads/journees/2253_fr.php

[14] Voir à ce sujet l'ouvrage collectif « *Les défis de la publication sur le web : hypertextes, cybertextes et méta –éditions* » sous la direction de JM Salaun et C Vandendorpe, Villeurbanne presse enssib, 2004.

connaissances (bases termino-ontologiques). Par conséquent, la structure des ressources, les fonctions et les terminologies employées reflètent l'héritage dont chaque institution-auteur se réclame. Même si l'hybridation est en marche (possibilités navigationnelles et d'annotation et d'extraction dans les ressources numériques obligent, introduction de relations conceptuelles issues d'autres types de ressources), certains dispositifs gardent encore, presque jalousement, des traces tangibles de leur domaine de provenance. Ceci est perceptible dans leur modèle conceptuel, dans la terminologie employée pour désigner leurs entités et les relations entre elles, et dans celles utilisées au niveau des fonctionnalités proposées par l'interface.

L'analyse fonctionnelle des ressources menée ici va nous conduire naturellement vers l'élaboration d'une deuxième grille d'évaluation, qui sera soumise aux usagers eux-mêmes (professionnels de l'information, journalistes, traducteurs utilisateurs novices). Ceux-ci devront évaluer l'utilité de ces ressources dans la recherche d'équivalences multilingues (fonctions utiles, manquantes) et dans l'appropriation des connaissances dans un contexte multilingue. Ils devraient également évaluer la charge cognitive que représentent les différents modèles de parcours proposés. Cette deuxième grille d'évaluation nous conduira à formuler des propositions en vue de la construction d'un dispositif intégré d'accès aux connaissances terminologiques dans un contexte multilingue permettant une meilleure prise en compte des attentes différenciées des usagers professionnels.

Bibliographie

BERTRAND, A. (1993) *Compréhension et Catégorisation dans une activité complexe : l'indexation de document scientifiques*. Thèse de Doctorat Nouveau Régime, Université de Toulouse-Le Mirail, Toulouse, 1993, 341p

DANCETTE J. (2003), *Les représentations lexico-sémantiques (RLS), moyen de structuration des connaissances dans les domaines spécialisés*, in actes colloque ISKO Grenoble 2003, pp 83-96

DENECKER C. (2002), *Les compétences documentaires : des processus mentaux à l'utilisation de l'information*, Villeurbanne : Presses de l'Enssib, 2002, 208p

DYENS O. (2004), *Le web et l'émergence d'une nouvelle structure*…., in « Les défis de la publication sur le web : hyperlectures, cybertextes et méta-éditions », coordonné par J.M. Salaun et C. Vandendorpe, Villeurbanne : éditions enssib, collection « Références », 2004, pp 205-215

EL HACHANI M. (2005), *Indexation des documents multilingues d'actualités incluant l'arabe : équivalence interlangues et gestion des connaissances chez les indexeurs*, Thèse de doctorat, Université Lumière Lyon, novembre 2005, 452p

GILCHRIST A. (2003), Thesauri, taxonomies and ontologies – an etymological note, *Journal of Documentation*, vol. 59(1), 2003, 7-18.

IBEKWE-SANJUAN F. (2006) , Clustering semantic relations for constructing and maintaining knowledge organization tools, *Journal of Documentation*, Emerald Publishing Group, vol. 62 (2), 2006, 229-250.

LARIVIERE, L. (1989), *Vers un produit unifié en terminologie et en documentation : le thésaurus terminologique,* in revue Méta, XXXIV, numéro 3, 1989, p.457

LYADRI R. (1997), *Problématique des équivalences sémantiques et de la traduction dans des dictionnaires arabe français*, Revue Méta, vol. 42, n°1, mars 1997, pp. 142-14

NASSE-KOLMAYER E. (1997), *Contribution à l'analyse des processus cognitifs mis en jeu dans l'interrogation d'une base de données documentaires*, Thèse de Doctorat en psychologie, Paris : Université Paris 5,1997, 335p

ORIGGI, G. (2004), *Pour une science humaine de l'internet*, in « *Les défis de la publication sur le web : hyperlectures, cybertextes et méta-éditions »*, coordonné par J.M. Salaun et C. Vandendorpe, Villeurbanne : éditions enssib, collection « Références », 2004, pp 219-243