

Education à la culture informationnelle multilingue : Outils et ressources pour les traducteurs

Jean-Paul Pinte

Docteur en Sciences de l'Information

**Enseignant - Chercheur au Laboratoire d'Ingénierie pédagogique de l'Institut
Catholique de Lille – France**

Mail : pinte.jp@gmail.com

Blog : <http://veillepedagogique.blog.lemonde.fr>

Introduction

Parmi les quelques questions vives qui animent les travaux de recherche autour du développement des TIC en éducation se pose aujourd'hui essentiellement celle de l'éducation à la culture informationnelle visant à la recherche pertinente d'informations et de ressources pédagogiques au travers des réseaux d'information numériques.

A l'heure du Web 2.0 et bientôt du Web 3.0, nos pratiques de recherche informationnelle oscillent encore trop souvent entre « sérendipité » et « zemblanité ». Il convient de répondre de manière urgente à la question du « Comment aller au-delà de ce que nous propose les moteurs de recherche traditionnels ? » en intégrant les concepts de Web invisible et de Web abyssal.

Aujourd'hui, dans la plupart des cas, on remarque que les recherches sur la toile se limitent le plus souvent à la langue de l'utilisateur alors que la dimension linguistique ouvre une chance supplémentaire de « cyberquête » pertinente qu'il convient d'intégrer dans le processus de culture informationnelle que l'on soit enseignant, étudiant, voire encore traducteur.

Quelles sont les motivations d'une éducation à la culture informationnelle, quels sont les outils cross-lingue et multilingue permettant d'aller de la simple requête sur un moteur de recherche aux agents d'automatisation de l'information en passant par la cartographie d'un concept ? Telles sont les questions auxquelles cet article tente de répondre.

Mots-clés: Education à la culture informationnelle, veille technologique, cross-lingue, veille multilingue, habiletés informationnelles, veille pédagogique

1 – Les langues oubliées dans l'Education à la culture informationnelle

Maîtriser l'information au XXI^e siècle est devenue indispensable pour quiconque souhaite survivre à une société dite de l'information, y trouver culturellement, économiquement et socialement sa place là où les médias, les moyens d'accéder à l'information et les ressources en ligne sont devenues chronophages et trop diversifiés.

Des organisations comme l'UNESCO vont même jusqu'à reconnaître cette « capacité à maîtriser l'information » comme une préoccupation éducative à l'échelon international. Ainsi, la proclamation d'Alexandrie en Egypte portant sur la maîtrise de l'information et l'apprentissage tout au long de la vie du 6 au 9 novembre 2005 nous rappelle l'histoire de cette écolière de dix ans qui a sauvé une centaine de vies grâce à sa maîtrise de l'information, alors qu'elle passait des vacances sur l'île thaïlandaise de Phuket au moment du tsunami de décembre 2004. Originaire d'Oxshott (Angleterre), Tilly Smith, qui avait étudié les tsunamis en cours de géographie deux semaines avant les vacances, a reconnu les signes avant-coureurs du phénomène et donné l'alerte. Grâce à sa capacité à utiliser les connaissances apprises, la plage a pu être évacuée et il n'y a eu aucune victime¹.

Spécialiste de la veille et de la culture informationnelle depuis 2000, j'ai aussi longtemps pensé qu'il était temps de mettre en place une éducation à l'information et ce, dès le plus jeune âge. J'ai donc débuté la recherche informationnelle avec mes étudiants à une époque et un temps où l'on se consacrait encore beaucoup plus à la découverte de la toile et à un émerveillement des internautes face aux capacités illimitées de ce fabuleux outil plutôt qu'à une démarche organisée de recherche informationnelle.

En mai 2003, Brigitte Juanals, dans un ouvrage intitulé « *La culture de l'information, du livre au numérique* » aux éditions Lavoisier repère les modèles culturels de connaissance proposés par les livres et les dispositifs informatisés, non seulement dans les supports et techniques, dans les modes de rassemblement, de classification et d'accès à l'information, mais aussi dans la représentation et la mise en scène des données. Le livre analyse aussi divers dispositifs de médiation aux informations, notamment la pratique éditoriale de l'*Encyclopædia Universalis* dans ses choix entre livres, disques compacts et sites en ligne. En portant un regard critique sur la mutation des accès à l'information, devenue dynamique, hypertextuelle et interactive,

¹ <http://portal.unesco.org/>

l'ouvrage évalue également dans quelle mesure elle peut aller dans le sens d'une démocratisation des connaissances.

Plus tard, en 2007, la contribution (p.25) d'Anne Lehmans de l'IUFM d'Aquitaine intitulé «*Eduquer à/par l'information*» est particulièrement intéressante et accessible. Elle est consacrée à l'«*information literacy*», que les francophones traduisent par «*alphabétisme informationnel*», «*culture de l'information*», «*intelligence informationnelle*», «*formation à l'information*».

Récemment, Alexandre Serres dans le dernier ouvrage de Fabrice Papy² a consacré un chapitre complet à cette problématique (Chapitre 5).

Même si le Colloque international de l'ERTÉ³ sur l'éducation à la culture informationnelle les 16, 17 et 18 octobre 2008 à Lille ouvrira le débat sur les enjeux et les représentations liés aux pratiques de la culture informationnelle, rien dans le programme ne semble cependant aujourd'hui porter une attention particulière à la pratique des langues dans le processus de recherche de l'information.

2 - La recherche de l'information : entre « sérendipité » et « zemblanité »

L'adaptation du fonctionnement de notre esprit selon le modèle de l'ordinateur (Computation) a, ces dernières années modifié nos modes d'accès aux connaissances.

La conception computationnelle décrite par Bruner⁴ (p.16) s'intéresse au traitement de l'information et cherche à savoir comment une information donnée sur le monde, information à la fois finie, codée et dépourvue d'ambiguïtés, va s'inscrire, être classée, mise en mémoire, assemblée, extraite et, plus généralement, maîtrisée par un outil computationnel.

Ce «*computationalisme*» est aujourd'hui opposé selon Bruner (p.18) au *culturalisme* qui cherche à réunir les acquis de la psychologie, de l'anthropologie, de la linguistique, de manière à reformuler un modèle de l'esprit et par conséquent une construction de la signification.

² Papy, Fabrice, «*Sciences de l'Information. Problématiques émergentes*». Londres, Paris : Hermès Lavoisier, 2008.

³ <http://ertecolloque.wordpress.com/>

⁴ Bruner.J, «*L'éducation, entrée dans la culture*», Ed.Retz, 2008

Si la machine apporte, il est certain une aide précieuse à la recherche d'information dans certains corpus de connaissances, il convient de signaler qu'il paraît encore difficile aujourd'hui de distinguer ces concepts dans les pratiques des internautes qui oscillent encore le plus souvent entre sérendipité et semblantité.

Selon une définition les plus courantes et notamment un billet sur ses origines paru dans le site de l'Urfist, la sérendipité serait « l'art de faire des découvertes heureuses, inattendues et utiles par hasard⁵ ».

La semblantité désignerait quant à elle « la faculté de faire de façon systématique des découvertes malheureuses, malchanceuses, attendues et n'apportant rien de nouveau⁶ ».

En général, les internautes oublient de rechercher les ressources « papier », les banques de données, les ressources ouvertes. Ils ignorent aussi les portails spécifiques, confondent les documents électroniques offerts par les bibliothèques et les ressources Internet en libre accès pour tous. Le concept de Web invisible leur reste étranger et ils ne savent pas citer les documents numériques ni les bibliographies.

Les repères critiques lui manquent aussi et Google et Wikipedia deviennent ses principales encyclopédies. Mais surtout ils ne pensent jamais à formuler leur recherche dans une autre langue.

3 - Une approche de la veille multilingue

Le principal objectif de la recherche d'information multilingue est de fournir à l'utilisateur qui ne serait pas familier avec une langue particulière, mais qui serait quand même intéressé à obtenir des documents dans une autre langue ou plusieurs autres langues.

Les principales tâches liées à la recherche d'information multilingue sont le filtrage, la sélection et le classement de documents qui pourraient être pertinents pour l'utilisateur.

⁵ http://urfistinfo.blogs.com/urfist_info/2005/04/srendipit_quand.html

⁶ <http://automatesintelligents.blog.lemonde.fr>

On peut noter avec l'expérience que certaines recherches menées dans ce cadre débouchent souvent sur des sites dont les résultats proposent à la grande surprise de l'internaute des résultats en plusieurs langues. L'utilisateur peut aujourd'hui faire cette démarche de manière moins hasardeuse avec LIRIX⁷, moteur de recherche multilingue développé sur la base du moteur Jakarta - Lucene chez XEROX. Ce même logiciel a pour particularité unique de pouvoir taper quelques mots en français en vue d'obtenir des documents à la fois en Français et en Anglais.

Mais aujourd'hui, même si tous ces outils évoluent, beaucoup de travail reste à faire pour obtenir le même niveau de performance de recherche obtenu avec les systèmes de recherche monolingues. La culture de la veille multilingue exige d'intégrer toutes les étapes du cycle d'une veille traditionnelle que sont : la détection des sources, leur collecte, leur capitalisation, diffusion et exploitation que seules certaines plateformes commencent à proposer aujourd'hui aux professionnels de la traduction.

Parmi les grands acteurs du marché on peut signaler aujourd'hui:

- Acetic
- Ayonis
- Clikya
- Grimmersoft
- Intellixir
- Knowings
- Lingway
- Mapstan
- Mathéo
- Noheto
- Pertimm
- Systran
- Synequa
- Thales
- Trivium
- Vecsys

⁷ <http://www.xrce.xerox.com/>

- .../...

4 - Typologie des outils de recherche multilingue

Les avancées de certains moteurs de recherche intégrant des modules de traduction automatique comme Google⁸ et Altavista avec Babel Fish⁹, voire encore l'outil de l'Education Nationale EUREKA¹⁰ pour ne pas tous les citer, ne font pas partie de la typologie décrite ci-dessous.

Il faut néanmoins signaler des avancées significatives sont prédites sur ces outils dans un proche avenir à l'instar de Google qui en dehors des dictionnaires intégrés à son moteur de recherche intègre de plus en plus les expressions courantes. Le futur centre de traduction de Google est très certainement destiné à révolutionner le monde de la traduction professionnelle à en croire le billet de Jean-Marie Le Ray¹¹ : *« Cet opérateur ira également rechercher les traductions dans les bitextes qui peupleront la mémoire de traduction universelle quotidiennement élaborée par une armée de traductrices et de traducteurs ».*

4-1 Les moteurs de recherche basés sur la traduction automatique

Ces outils partent d'une requête formulée par l'utilisateur ou encore d'un corpus de documents. La précision est moindre que la collection de documents qui contient un contexte d'information nettement plus important et c'est donc la traduction de la requête qui est retenue lorsque l'on parle de recherche d'information multilingue sur ces outils.

Cette méthode est en cours de consolidation et les résultats qu'elle donne ne sont pas très satisfaisants, car elle se base sur la traduction automatique dont les résultats ne sont pas toujours probants eux aussi, surtout lorsque les requêtes sont courtes. En effet plus les requêtes sont courtes, moins les résultats seront bons.

4-2 Les moteurs de recherche basés sur un corpus d'apprentissage

Cette méthode s'appuie, pour la traduction des requêtes sur un thésaurus créé (grâce à un algorithme) à partir d'un corpus d'apprentissage aligné au niveau de la phrase pour trouver des cooccurrences de termes en contextes (Huthins, 2005).

Les résultats de cette méthode sont généralement assez satisfaisants mais ont l'inconvénient d'être limités. Comme pour toute méthode statistique, on constate un arrêt de la progression des résultats à un certain seuil.

⁸ http://translate.google.com/translate_s?hl=fr

⁹ <http://fr.babelfish.yahoo.com/>

¹⁰ <http://www.educnet.education.fr/dossier/rechercher/traduction1.htm>

¹¹ <http://adscriptum.blogspot.com/2008/08/google-et-la-traduction.html>

4 - 3 Les moteurs de recherche basés sur des dictionnaires de reformulation

Cette méthode appelée en anglais « *machine-readable dictionaries-based method* » se base sur l'expansion de la requête.

Cela consiste à formuler la requête autrement en remplaçant les mots qui la composent par des variantes afin de récupérer des documents pertinents dans lesquels les termes saisis ne sont pas toujours présents.

Cette reformulation se fait à l'aide de dictionnaires monolingues permettant la reformulation dans une même langue (synonymes, antonymes, etc.) et les dictionnaires bilingues qui permettent la reformulation dans des langues différentes (Aljlal et *al.* 2001).

4 - 4 Les moteurs utilisant une langue pivot

Cette méthode consiste à extraire la sémantique des textes de la langue source grâce à un langage pivot.

Il s'agit d'un langage unifié permettant de représenter la sémantique des différentes langues et basé sur le concept des graphes représentant les phrases.

Cette méthode est la moins consolidée de toutes celles que nous avons citées et peu de résultats ont été constatés (Hahn et *al.* 2004)

5 – Vers une typologie des outils

5-1 Knowings¹² multilinguisme : traduction automatique des requêtes et interprète virtuel

Global Finder propos un ensemble d'outils dont la génération automatique de requêtes multilingues rendue possible tout en intégrant le processus et le paramétrage complet d'une veille dite professionnelle.

Chaque requête peut être automatiquement traduite en cinq langues (anglais, français, espagnol, allemand et italien) pour pouvoir adresser simultanément des sources de langues différentes. L'adaptation des dictionnaires de traduction automatique à la terminologie métier de l'entreprise y est rendue possible

Une assistance à la compréhension des résultats grâce à un véritable interprète virtuel permettant de traduire « à la volée » des mots et des expressions au sein des résultats est comprise dans l'outil et cette traduction prend en compte le contexte sémantique de la phrase et s'appuie sur le vocabulaire « métier » éventuellement paramétré.

5-2 Pertinence Mining¹³ : Résumé automatique de textes et analyse linguistique du discours

Pertinence Summarizer effectue par exemple un traitement linguistique d'un document et évalue la pertinence des phrases. Le procédé prend en compte non seulement des marqueurs linguistiques généraux et/ou spécialisés selon la nature du document analysé mais également les mots-clés de l'utilisateur et éventuellement des bases terminologiques pour renforcer la pertinence des phrases retenues.

Pertinence Summarizer prend en charge une dizaine de langues parmi lesquelles figurent une grande partie des langues européennes, des langues asiatiques (japonais, chinois, coréen), l'arabe et le russe.

Les formats de documents supportés par *Pertinence Summarizer* sont : HTML, PDF, DOC, RTF, TXT et d'autres formats tels que PPT, XLS selon les plates-formes.

Pertinence Summarizer est accessible avec un navigateur web. Disponible en version serveur ou monoposte *Pertinence Summarizer* peut être installé sur tout poste équipé d'une machine virtuelle Java. Une API Java et une API Web Service permettent également d'intégrer les fonctionnalités de *Pertinence Summarizer* dans vos applications. (<http://www.pertinence-mining.com/PSummarizer.pdf>)

5-3 Lingway¹⁴: L'ingénierie linguistique

Des progrès considérables ont été enregistrés grâce à Lingway ces dernières années dans le domaine du TAL (Traitement Automatique du Langage Naturel - Natural Language Processing (NLP) en anglais) ou de l'Ingénierie Linguistique, son autre appellation. Et ceci à

¹² <http://www.knowings.com/>

¹³ <http://www.pertinence.net/>

¹⁴ <http://www.lingway.com/content/view/20/60/lang,fr/>

partir du moment où l'on a pu décomposer le problème, trop vaste pour être abordé globalement, en plusieurs sous-problèmes, correspondant eux-mêmes à plusieurs sous-tâches articulées entre elles. A noter, que cette décomposition est faite à la fois au niveau des ressources linguistiques (dictionnaires et règles de grammaire) et au niveau des algorithmes (analyseurs, transducteurs, ...).

Le cœur de la technologie Lingway KM est basé sur ces différents niveaux d'analyse, pour lesquels autant de modules spécialisés sont utilisés.

Aujourd'hui on distingue principalement 5 niveaux d'analyse sur Lingway :

- Le niveau morphologique : identification des mots d'une phrase,
- le niveau syntaxique : identification des constituants et des fonctions d'une phrase,
- le niveau sémantique : identification du sens des mots et de la structure logique d'une phrase,
- le niveau du texte : identification des relations entre les phrases et de la structure d'un texte,
- le niveau du corpus : identification de la structure d'un ensemble de textes.

5 - 4 Sinequa ou la pertinence de la navigation

Sinequa CS¹⁵ utilise la combinaison de sa technologie brevetée d'analyse sémantique et linguistique, avec l'analyse statistique, structurée et morphosyntaxique pour retrouver, organiser et naviguer à travers les éléments de connaissance les plus pertinents par rapport à une recherche.

Pour que l'utilisateur puisse naviguer et atteindre rapidement le ou les éléments les plus pertinents, SINEQUA CS organise l'information grâce à une génération dynamique de concepts et de catégories métier.

Cela fournit à l'utilisateur une vue complète et pertinente des réponses possibles, lui permettant de prendre une décision appropriée plus rapidement.

5 - 5 Kaliwatch, plateforme intelligente de veille intégrée

Kaliwatch Server¹⁶ est une solution pertinente de la Compagnie Thales.

Robuste et éprouvée, elle associe classement automatisé et moteur de recherche sémantique, pour vous permettre de trouver les moyens d'atteindre vos objectifs et d'améliorer la productivité.

Complète et intégrée, elle surveille, collecte, organise, analyse et diffuse l'information interne et externe aux entreprises. Ses principales fonctionnalités sont les suivantes :

¹⁵ <http://www.sinequa.com/solutions.html>

¹⁶ <http://www.arisem.com/>

- Collecte Multicanal pour accéder à tout type d'information :

- Bases de données
- Internet visible et invisible grâce à des agents intelligents et furtifs
- Systèmes de fichiers

- Traitements automatisés pour garantir une excellente pertinence des résultats :

- Analyse sémantique crosslingue
- Classement automatique
- Catégorisation
- Extraction d'entités nommées

- Diffusion précise à partir d'un portail de consultation, de recherche et de partage des informations :

- Accès personnalisé et sécurisé à l'information
- Recherche sémantique, full text, de catégories, de profils d'experts, etc.
- Edition d'actualités, de bookmarks
- Participation à des forums

Kaliwatch Server dispose de plusieurs modules optionnels à haute valeur ajoutée, pour lesquels vous pouvez à votre aise consulter les fiches produit sur la présente page.

5 - 6 Pertimm : Accéder rapidement à une information pertinente

La pertinence d'une réponse dépend de sa formulation. Fondé sur la linguistique et la sémantique, **Pertimm** permet aux utilisateurs de détailler tous les éléments de leur requête : plus ils entrent de données, meilleur est leur résultat. Voici ses principales fonctionnalités :

- Gestion dynamique des synonymes et des thesauri

Si vous disposez déjà d'une liste de synonymes ou d'acronymes, d'un ou plusieurs thesauri très importants pour votre métier, vous avez la possibilité de les ajouter à Pertimm, et de les modifier quand vous le souhaitez : nul besoin de tout ré-analyser.

- Gestion complète des entités nommées

Vous pouvez intégrer dans Pertimm des listes de noms (de produits, de personnes, de sociétés, etc.) : leur présence vous sera indiquée dynamiquement dans les résultats de chaque requête.

- Navigation sémantique dans les contenus

Vous pouvez intégrer à Pertimm des concepts, que nous appelons Pertimmiseurs : ils seront retrouvés dans la liste des résultats. Grâce aux Pertimmiseurs, toutes les fonctions dont vous avez besoin deviennent dynamiques : Catégorisation, Classement, Ontologies...

- Fabrication automatique de résumés ou "aperçus sémantiques"

Pour chaque document, Pertimm affiche un « aperçu » qui permet de le différencier par rapport aux autres dans la liste.

- Tri des résultats par Pertinence ou par date

Par défaut, les résultats sont triés par pertinence par rapport à la requête, mais vous pouvez aussi décider de les trier par date.

- Visualisation des contenus avec mise en surbrillance

Pertimm vous montre au sein des documents les termes correspondants à votre requête, les déclinaisons, ou encore les synonymes.

5 - 7 Mtranslator¹⁷ ou la plateforme interactive multilingue

Cette plateforme offre une manière plus facile de rechercher sur le Web dans beaucoup de langues. La recherche multilingue d'Imtranslator fournit l'environnement de recherche soutenu par correcteur orthographique multilingue, traducteur, clavier virtuel, et décodeur russe

5 - 8 Ouverture du concept de cartographie à la recherche multilingue

La cartographie d'informations est un domaine en pleine expansion. Dans un contexte de surabondance informationnelle, cette méthode de représentation et de visualisation présente des perspectives prometteuses et suggèrent des applications extrêmement diversifiées dans une multitude de secteurs.

La cartographie informationnelle telle que nous l'entendons aujourd'hui est différente de celle la cartographie de données sur une base géographique ou géolocalisation. On évoque plus ici la cartographie de données abstraites (idées, concepts, unités d'information), c'est-à-dire ce que l'on appelle plus communément la cartographie d'informations et de connaissances.

¹⁷ <http://imtranslator.com/#>

La visualisation d'informations est « l'utilisation informatisée de représentations visuelles interactives de données abstraites afin d'amplifier la cognition¹⁸ ».

Des plateformes comme Clusty, Kartoo, Mapstan, Ujiko, Touchgraph ont développé des accès multilingues sur la toile dans leur module de recherche avancée. Le site Mapdream.com propose aux personnes intéressées une cartographie des outils de cartographie recensés à ce jour dans différents domaines de recherche.

Pour la plupart ils intègrent aussi maintenant la dimension du Web invisible permettant d'accéder à des sites internes, des bases de données, des portails spécifiques, bibliothèques, etc.

5 - 9 La mobilité dans le décor du multilinguisme

Conscients que le téléphone mobile et ses évolutions dans un avenir proche sont voués à un développement sans précédent dans la société de l'information, des sociétés intègrent aujourd'hui la dimension « mobilité » dans le développement de leurs produits.

Dictionnaire de définitions, de synonymes, de traduction, les dictionnaires **Ultralingua** contiennent jusqu'à 300.000 entrées, y compris des milliers d'expressions argotiques et de termes techniques. Ils offrent en outre la conjugaison des verbes et une touche active permettant d'accéder aux dictionnaires avec un simple clic. La gamme comprend des dictionnaires de définitions en anglais et en français, avec de multiples dictionnaires bilingues pour l'anglais, le français, l'espagnol, l'italien, le portugais, le norvégien, et bien d'autres langues encore.

Les dictionnaires Ultralingua sont disponibles pour Macintosh OS X, Microsoft Windows, Palm OS, et Pocket PC (Windows Mobile).

Autre exemple, celui de **JLearnItME**¹⁹ est un traducteur multilingue pour téléphones portables acceptant Java avec 15 langues.

Le même dictionnaire peut traduire de n'importe quelle langue vers n'importe quelle langue (ex: anglais->français et français->anglais). L'Interface utilisateur est disponible en anglais, français, néerlandais, allemand, italien et tchèque et, de plus la possibilité de créer votre propre dictionnaire vous est offerte.

¹⁸ Card S. K., Mackinlay J. D., Shneiderman B., Information Visualization : Using Vision to Think, p.10-11

¹⁹ http://www.jlearnit.com/jlearnitme/index_fr.html

Notons enfin le cas de **InterKey** permet d'entrer des données en langue nationale dans les téléphones portables à Windows Smartphone (smartphones) via le clavier: le russe, le grec, etc. (support complet de la langue nationale avec l'interface de traduction, et est disponible comme une version particulière).

6 - Conclusion

Trouver l'information à l'aide de moteurs de recherche traditionnels et d'automates sur une toile où celle-ci devient pléthorique demeure, nous l'avons vue précédemment, encore un vœu pieu.

Le développement de la recherche multilingue se fait donc de plus en plus sentir pour tous les acteurs de la société, et principalement les traducteurs, dont le rôle primordial ira de plus en plus dans le sens d'une veille multi-lingue et cross-lingue.

Il faut aussi tenter de sortir au plus vite de l'« alphanétisation » qui voudrait que l'on s'exprime et que l'on apprenne en toute autonomie avec les outils existants sur le Web.

Il faut en réinventer d'autres dans notre société, former et éduquer ses principaux acteurs à la culture informationnelle multilingue tout en leur faisant comprendre les enjeux et les implications personnelles et professionnelles des évolutions technologiques et sociologiques d'une époque dite sans frontières mais où l'importance des langues n'a jamais été aussi mise en avant.