

Quatre actions de recherche en réseau sont financées en 2009 dans le cadre du réseau LTT.

Reflets culturels de la métaphore de spécialité d'une culture à l'autre

*Responsable : Teresa LINO, Université Nova de Lisboa, Portugal (Budget : année 1 : 8500 €),
pays des principaux partenaires : Portugal, France, Sénégal, République du Cap-Vert*

L'étude de la métaphore participe en particulier à la préservation de la diversité linguistique et des cultures que les différentes langues véhiculent. De plus, la métaphore de spécialité est le résultat d'une construction chargée de valeurs culturelles. En d'autres termes, la métaphore présente des significations et des associations qui ont cours dans un cadre culturel. Son interprétation et sa production renvoient à des analogies culturellement variables. D'une culture à l'autre, il existe de nombreuses corrélations qui déterminent et expliquent les fondements de ces structures métaphoriques. En effet, notre propre vision du monde détermine notre façon de classer, d'ordonner et de catégoriser ce que l'on perçoit.

Cette recherche a comme thématique centrale la métaphore terminologique en langues de spécialité, en particulier, dans le domaine de la médecine tropicale ; mais le projet a comme objectif la constitution de ressources mono et multilingues pour le recours dans le domaine professionnel et didactique. Cette recherche développera les rapports entre langue française et langues partenaires : le portugais, le wolof et le capverdien.

Dans ce projet, nous faisons collaborer différentes équipes qui regroupent des compétences dans les différentes langues étudiées : des compétences en terminologie et des compétences en traitement automatique des langues, des compétences en technologies nouvelles pour l'analyse et l'exploitation et la diffusion des corpus multilingues.

Ces dernières années, la métaphore a fait l'objet de nombreux travaux en linguistique, mais aussi en sciences cognitives, en neurolinguistique, en anthropologie. Ce travail exige une nouvelle méthode de travail basée sur une approche culturelle de la terminologie, approche qui permettra de commencer à poser des jalons théoriques.

Teresa LINO
Universidade Nova de Lisboa

Projet *MotAMot* : Élaboration d'un système lexical multilingue par le biais de la construction de dictionnaires bilingues ciblés sur les langues peu informatisées d'Asie du Sud-Est

Responsable : Mathieu MANGEOT-NAGATA, Université Joseph Fourier, Grenoble, France : (Budget : année 1 : 4535 €), pays des principaux partenaires : France, Vietnam, Cambodge

Le projet *MotAMot* a pour but de construire un dictionnaire multilingue entre le français et des langues d'Asie du Sud-Est, en priorité le vietnamien et le khmer. Ce projet est le fruit d'une coopération entre les laboratoires GETALP-LIG de Grenoble (France), MICA de Hanoï (Vietnam) et ITC de Phnom Penh (Cambodge).

La macrostructure du dictionnaire est basée sur une structure pivot : chaque sens de mot de chaque langue est relié au volume pivot par un lien interlingue. De cette façon, l'établissement de deux liens sens langue A > sens langue B et sens langue B > sens langue C a pour conséquence l'obtention d'un lien sens langue A > sens langue C. La microstructure est basée sur la lexicographie explicative et combinatoire avec un niveau de description très fin et très détaillé. Chaque article décrit un vocable. Les vocables sont ensuite divisés en sens de mots ou lexies. Chaque lexie est décrite avec une formule sémantique, sorte de définition formelle, un lien interlingue vers le volume pivot, une liste de fonctions lexicales et des exemples.

La construction du dictionnaire sera réalisée en ligne sur la plate-forme *Jibiki* de gestion de ressources lexicales par des contributeurs bénévoles. Un niveau de qualité est attribué aux données : d'une étoile pour une qualité brouillon à cinq étoiles pour une qualité certifiée. De même, les contributeurs auront un niveau de compétence défini : d'une étoile pour les débutants à cinq étoiles pour les experts. Si un contributeur révise un article d'un niveau inférieur, celui-ci prend le niveau du contributeur. D'autre part, l'interface d'édition d'un article s'adaptera automatiquement au niveau du contributeur. Par exemple, pour les débutants, la formule sémantique sera remplacée par une simple glose. Ils n'auront pas non plus de fonctions lexicales à renseigner. Pour les experts par contre, ils auront une vue détaillée de toutes les informations disponibles.

La méthodologie d'élaboration de la ressource est la suivante. Tout d'abord, constituer des communautés de personnes intéressées pour contribuer. Nous utiliserons pour cela les jeux sérieux lexicaux de type *JeuxDeMots*. Les ressources ainsi constituées seront par la suite intégrées dans le dictionnaire. Ensuite, récupérer des données existantes. Nous avons plusieurs ressources français-vietnamien déjà informatisées. Pour ce qui est du khmer, nous prévoyons d'informatiser un dictionnaire existant français-khmer. Enfin, proposer la contribution directe en ligne.

Le projet est déjà relativement avancé. Une version beta de la plate-forme est déjà disponible. Il reste à la tester et à rajouter les aménagements spécifiques au projet. Pour les *JeuxDeMots*, la version française, créée dans le cadre d'un autre projet, est disponible depuis deux ans. La version vietnamienne vient d'être traduite et mise en ligne pour test. Nous prévoyons de démarrer l'informatisation du dictionnaire français-khmer au printemps 2010 avec la venue d'un informaticien cambodgien.

Mathieu MANGEOT-NAGATA
Université Joseph Fourier

Base de lexicographie linguistique bilingue : traduction, corpus parallèles, terminologie

Responsable : Salah MEJRI, Université Paris 13, France : (Budget : année 1 : 8100 €), pays des principaux partenaires : France, Tunisie

La création de la base de données Métalangue répond à trois objectifs : (i) la constitution d'un observatoire de la terminologie des Sciences du langage, de son évolution, de sa variation en diachronie et en synchronie, en fonction des courants, des écoles et des orientations propres aux auteurs ; (ii) descriptive par définition, cette base constitue un outil destiné à la formation des (jeunes) chercheurs à qui elle fournit un panorama analytique dans lequel les unités sont contextualisées, leur signification décrite et, si nécessaire, rapportée aux prérequis de l'appareil théorique dans lequel elles ont vu jour ou bien sont majoritairement employées ; (iii) fournir une aide à la traduction et à l'équipement métalinguistique des langues du Sud, en premier lieu l'arabe. La base de données *Métalangue* comporte actuellement près de 7 000 fiches concernant le français et à peu près 2 000 portant sur l'arabe. Répondant à un objectif de formation, le corpus d'observation est constitué massivement de manuels et d'ouvrages connaissant une forte diffusion dans les Sciences du langage, ainsi que de textes français traduits en arabe. Elle est gérée par une équipe internationale établie en France et en Tunisie, copilotée par le laboratoire LDI (Lexiques, Dictionnaires, Informatique – UMR 7187, Université de Paris XIII) et le laboratoire international LTA (Langues, Traduction, Apprentissage ; universités de Paris XIII, de Tunis et de Sousse). Dans le courant de l'année 2010, la base *Métalangue* sera accessible librement par Internet aux chercheurs et institutions.

Gérard PETIT
LDI – UMR 7187, Paris XIII

Bases de données pour l'étude grammaticale et lexicale des langues dans une visée multilingue

Responsable : Patrice POGNAN, INALCO, France (Budget : année 1 : 8290 €), pays des principaux partenaires : France, Albanie, Maroc

Ce projet a pour but la réalisation de « chantiers de données » pour combler un besoin très fort de données précises et exhaustives sur les langues, mais aussi d'outils permettant la création, la gestion et la consultation de ces données tout en donnant l'opportunité de nombreuses applications en traitement automatique des langues, en création de dictionnaires, de lexiques et de manuels, ainsi que pour l'apprentissage d'un groupe de langues ou d'une langue dans son cadre de référence (système linguistique du groupe ou de la famille de langues), ce qui constitue aussi un support à l'intercompréhension.

Le principal outil visé est un prototype de bases de données, multifonctionnel et multilingue ayant le français en langue pivot. Cet outil doit être en mesure de consigner les données de toute langue indo-européenne, de toute langue chamito-sémitique avec une ouverture vers les langues agglutinantes, principalement ouralo-altaïques et finno-ougriennes. Cet outil donne des modèles particuliers par simple duplication avec d'éventuelles modifications mineures. Il est accompagné de la constitution de *corpora*.

Nous ferons état, fin 2009, de deux grands axes de recherche :

► sur les langues indo-européennes occidentales par l'étude de l'albanais et des langues slaves de l'Ouest, actuellement tchèque et slovaque (avec le slovène comme référent des langues slaves du Sud, étude soutenue par un PHC Proteus), mettant en relief l'étude du système linguistique tant synchronique que diachronique. Une composante de génération flexionnelle interne à ces bases de données est actuellement en cours de réalisation ;
► sur le berbère, langue chamito-sémitique, où le grand « dictionnaire raisonné berbère-français. Parlers du Maroc » de Miloud TAÏFI (objet d'un soutien par une action intégrée et un FSP France-Maghreb) devient un corpus à partir duquel on construit automatiquement une base de données suivant le modèle du prototype. Ce corpus et cette base de données, en voie d'achèvement, permettront la réalisation de nombreux autres corpora : nous possédons déjà « l'exemplier TAÏFI », c'est-à-dire un ensemble de 13 500 exemples berbères authentiques traduits en français et il est possible aussi d'extraire de la masse lexicale la matière pour la confection de paradigmes de synonymes et de locutions verbales. Par ailleurs, l'inversion du dictionnaire dans le sens « significations françaises-correspondants berbères » fournira la matière lexicale pré-organisée nécessaire à la construction d'un dictionnaire français-berbère.

Patrice POGNAN et Miloud TAÏFI
INALCO

Extrait de *Lettre des réseaux de chercheurs sur les cultures, les langues et les littératures*, n°6 novembre 2009