

RAPPORT DE FIN DE PROJET

Nom du programme (ou du pays) : Polonium

Titre du projet : Description morphologique de noms propres polonais pour applications multilingues

Nom du responsable français : Agata CHROBOT épouse SAVARY

I. Partenaires :

	France	Pays partenaire
Chef de projet : Nom : Fonction :	Agata CHROBOT ép. SAVARY Maître de conférences	Marcin WOLIŃSKI Chargé de recherche
Intitulé du laboratoire (mentionner le n° d'unité le cas échéant) :	Laboratoire d'Informatique EA 2101	Instytut Podstaw Informatyki (IPIPAN) NIP : 525-00-94-01 REGON : 000326635
Institution de rattachement (pour les Unités Mixtes, lister tous les partenaires) :	Université François Rabelais de Tours	Polska Akademia Nauk (Académie Polonaise de Sciences)
Ville :	Tours et Blois	Varsovie

II. Résultats scientifiques obtenus :

Décrire, en 30 lignes maximum, dans le cadre ci-dessous, les principaux résultats scientifiques obtenus à l'issue du projet.

Dans sa globalité, les résultats obtenus sont ceux annoncés dans le dossier de candidature du projet:

- 1) L'élaboration d'un formalisme pour la description morphologique des noms propres MONOLEXICAUX (inspiré du formalisme créé par [Przepiórkowski, Wolinski 2003] pour les noms communs du polonais). Ce formalisme est conforme à celui proposé pour les mots communs monolexicaux et consiste en une jeux d'étiquettes morpho-syntaxiques.
- 2) Développement d'outils informatiques d'aide à la description et traitement morphologique des noms propres MONOLEXICAUX (à la base du système morphologique MORFEUSZ [Woliński, 2006]). Ce système, effectuant initialement une analyse morphologique du polonais, a été élargi, spécifiquement pour ce projet, par un module de génération de formes fléchies de mots monolexicaux.
- 3) Création d'un dictionnaire électronique morphologique contenant des noms propres monolexicaux polonais et des noms communs faisant partie des noms propres polylexicaux (e.g. "ogród" et "luksemburski" dans "Ogród Luksemburski"). Les composants monolexicaux font partie du large dictionnaire électronique du polonais récemment édité.
- 4) Elaboration d'un formalisme pour la description des noms propres POLYLEXICAUX (inspiré du système MULTIFLEX proposé par [Savary, 2005] pour les noms communs). Ce formalisme a été perfectionné suite à des observations sur certaines propriétés des noms propres. L'amélioration majeure consiste en l'introduction des descriptions imbriquées (un nom propre polylexical qui contient un autre nom polylexical). Le système MULTIFLEX, développé précédemment dans le cadre du système multiligne Unitex, a été amélioré au niveau de son interopérabilité, afin de collaborer avec d'autres modules morphologiques pour les mots monolexicaux. Une interface avec le système Morfeusz, permet le traitement du polonais dans un modèle morphologique autre que celui de Unitex. La modularité obtenue pour MULTIFLEX grâce à ce projet permet désormais son interfaçage avec tout autre module morphologique pour toute autre langue, sous réserve de certaines contraintes d'interface.
- 5) Développement d'outils informatiques d'aide à la description morphologique des noms propres POLYLEXICAUX. Un développement important, en lien avec le projet européen LUNA, a permis la création d'un outil d'aide au codage de noms propres, où les systèmes MORFEUSZ et MULTIFLEX collaborent, et où la description morphologique devient automatisée.
- 6) Création d'un prototype de dictionnaire électronique morphologique des noms propres POLYLEXICAUX polonais. Actuellement ce prototype porte sur des représentants de quelques dizaines de classes morphologiques. Le travail du codage sera poursuivi en 2009.

Ces résultats ont-ils donné lieu à une (des) publication(s) dans des revues ou communication(s) dans des colloques ? Si oui, reporter ci-dessous les titres des travaux et les références correspondantes (veiller à mentionner le lieu des colloques).

Agata Savary, Joanna Rabięga-Wisniewska, Marcin Wolinski, "Inflection of Polish Multi-Word Proper Names with Morfeusz and Multiflex", à paraître dans un ouvrage collectif chez Springer (30 pages)

Agata Savary, Multiflex: an Interoperable Multilingual Tool for Multi-word Units, en cours de soumission (8 pages)

Un article en soumission pour "Urban place names 2009", Helsinki 13-16 août 2009

4 communications à des séminaires internes des deux instituts

Ces résultats ont-ils donné lieu à des retombées socio-économiques (dépôt de brevet, valorisation, partenariat avec des collectivités locales, etc.) ? Si oui, les mentionner ci-dessous (titre du brevet et sa référence, partenaires impliqués, etc.).

Contact préliminaire avec la société ESRI, spécialisée en systèmes d'informations géographiques, sur l'amélioration de processus de géocodage à la base des connaissances linguistiques des noms propres.

III. Formation par la recherche :

Mentionner ci-dessous les thèses soutenues ou débutées – éventuellement en cotutelle – dans le cadre du projet, en précisant les noms des doctorants et les sujets de thèse.

Initiation à la recherche d'un étudiant en master informatique, Piotr Sikora, auteur du logiciel d'aide au codage des noms propres.

IV. Ouverture à l'international :

Le projet a-t-il donné lieu à une ouverture vers d'autres partenaires ou institutions européens ou internationaux ? La participation à un programme de recherche européen ou international est-elle envisagée ?

Le présent projet a été, en sa deuxième année, rapproché du projet européen LUNA (IST-033549). Une partie des fonds disponibles pour LUNA a permis le développement de l'outil d'aide au codage des noms propres, ainsi que le codage lui-même qui a débuté en décembre 2008 et sera poursuivi jusqu'en septembre 2009.

Les acquis du système Multiflex en ce qui concerne l'extension du formalisme (structures imbriquées) seront intégrés dans un outil de création et gestion de ressources linguistiques WS2LR développé à l'Université de Belgrade, avec lequel nous avons eu un projet EGIDE PHC "Pavle Savic" en 2004-2005.

L'évolution du logiciel Multiflex a été l'une des motivations pour le dépôt d'un nouveau projet EGIDE PHC avec la Bulgarie.

V. Utilisation des moyens :

Déplacements	En France	Dans le pays partenaire
Nom, grade, âge du (des) chercheur(s) :	Marcin WOLIŃSKI, Chargé de recherche, 35 ans	Agata SAVARY (Chrobot), Maître de conférences, 38 ans
Date de la mission (début, fin) :	12-15 décembre 2007	14-23 avril 2007
Nom, grade, âge du (des) chercheur(s) :	Joanna RABIEGA-WIŚNIEWSKA, chercheuse contractuelle, 32 ans	Béatrice BOUCHOU, Maître de conférences, 42 ans
Date de la mission (début, fin) :	12-15 décembre 2007	14-18 avril 2007
Nom, grade, âge du (des) chercheur(s) :	Marcin WOLIŃSKI, Chargé de recherche, 36 ans	Denis MAUREL, Professeur, 51 ans
Date de la mission (début, fin) :	6-12 juillet 2008	14-18 avril 2007
Nom, grade, âge du (des) chercheur(s) :	Joanna RABIEGA-WIŚNIEWSKA, docteur, chercheuse contractuelle, 33 ans	Agata SAVARY (Chrobot), Maître de conférences, 39 ans
Date de la mission (début, fin) :	6-12 juillet 2008	1-12 mai 2008
Nom, grade, âge du (des) chercheur(s) :	Małgorzata MARCINIAK, Chargée de recherche, 46 ans	Agata SAVARY (Chrobot), Maître de conférences, 39 ans
Date de la mission (début, fin) :	6-11 juillet 2008	15-20 novembre 2008
Nom, grade, âge du (des) chercheur(s) :		Nathalie FRIBURGER, Maître de conférences, 35 ans
Date de la mission (début, fin) :		15-20 novembre 2008

VI. Commentaires et suggestions, difficultés rencontrées, etc. :

Nous pensons que le projet a bien joué son rôle, i.e. celui de l'initiation d'une collaboration bilatérale à poursuivre dans un cadre élargi.

Les fonds attribués pour le projet correspondaient assez bien aux besoins effectifs.

L'organisation financière du côté du ministère polonais laisse à désirer, car le laboratoire IPIPAN n'a effectivement reçu aucun moyen financier supplémentaire du fait de participer au projet. Le ministère polonais a simplement donné l'autorisation au laboratoire de flécher une partie du financement habituel pour ce projet. Ce fait joue en défaveur de futures collaborations de ce type.