TODIRASCU Amalia

Corps et grade : Professeur des Universités

Coordonnées : Faculté des Lettres, 22 rue René Descartes, 67084 Strasbourg cedex
Bureau : Patio, Bâtiment 4, 4408
Tel : 03 68 85 65 86
e-mail : todiras@unistra.fr

CURRICULUM VITAE

FORMATION ET DIPLOMES

- **HDR en Informatique**, Université de Tours, soutenue le 15 novembre 2013. Jury: M. Jean-Yves Antoine, Université de Tours (rapporteur), M. Denis Maurel, Université de Tours (garant), M.Laurent Romary INRIA (examinateur), Mme Isabelle Tellier, Université Paris 3 (rapporteur), M. Eric Wehrli, Université de Génève (rapporteur). Titre du mémoire HDR : « Contributions au Traitement automatique des langues : des corpus aux systèmes d'extraction d'information »
- **Doctorat en Informatique,** Université Louis Pasteur Strasbourg I, obtenu le 22 Mars 2001
  Titre : « Indexation sémantique pour les systèmes de recherche documentaire » (en anglais)
  sous la direction de :
  -Bernard Keith, Professeur à l'Université Louis Pasteur Strasbourg I
  -Dan Gâlea, Professeur à l'Université Technique "Gheorghe Asachi" de Iasi,
  Roumanie (thèse en co-tutelle 6 mois/an dans chaque université)
- **Diplôme équivalente DEA en Informatique**, Juin 1993
  Faculté d'Informatique, Université "Al. I. Cuza" de Iasi, Roumanie

EXPERIENCE PROFESSIONNELLE

09/2016 – présent Professeur des Universités, Faculté des Lettres, Université de Strasbourg
09/2004-08/2016 Maître de conférences, Département d'Informatique, UFR LSHA, Université de Strasbourg
09/2002-09/2004 Maître de conférences, FRE CNRS 2732 - Institut des Sciences et Technologies de l'Information de Troyes (ISTIT), Université de technologie de Troyes
06/2002-08/2002 Ingénieur expert, INRIA Lorraine (Institut National de Recherche en Informatique et Automatique).
09/2001-05/2002 Chercheur post-doctorant, LORIA, INRIA Lorraine (Institut National de Recherche en Informatique et Automatique).
09/1996-09/2001 ATER et Chargé de cours, Faculté d'Informatique, Université "Al. I. Cuza" Iasi, Roumanie
05/1994-05/1996 Assistant chercheur, Institut d'Informatique Théorique, Académie Roumaine - filiale de Iasi, Roumanie

ACTIVITES DE RECHERCHE

Spécialité : Traitement automatique des langues, extraction d'information, création des ressources électroniques pour la traduction automatique, annotation sémantique et discursive, moteurs de recherche thématiques, systèmes de résolution des chaînes de référence

Publications (de 2011 à 2016)

Livres :

Contributions à des ouvrages collectifs :

Todiraşcu, Amalia, Navlea, Mirabela, (2010), « CAP: A Hierarchical Lexical Function Related to Proper Nouns: The Case of Romanian and French », in Dan Tufiş and Corina Forăscu (eds.): *Multilinguality and Interoperability in Language Processing with Emphasis on Romanian*, Romanian Academy Publishing House, Bucharest, pp. 317-330, ISBN 978-973-27-1972-5.

Articles

Articles parus dans des revues à comité de lecture :

1) Longo, Laurence, TODIRASCU, Amalia. (2014) Vers une typologie des chaînes de référence dans des textes administratifs et juridiques, F. Landragin et C.Schnedecker (éds), numéro spécial "Les chaînes de

références" de la revue Langages, N. 195, 2014/3, ISBN 9782200929381, ISSN 0458-726X, septembre 2014, pp. 79-98.

2) TODIRASCU, Amalia, GRASS, Thierry, NAVLEA, Mirabela, LONGO, Laurence (2014), La relation lexicale «Chef» : une approche translingue français-anglais-allemand, *META: le journal des traducteurs*, 59(2), août 2014, ISSN : 0026-0452 (sous presse)

3) Navlea, Mirabela, Todiraşcu, Amalia (2013) A Hybrid Word Alignment System for Statistical Machine Translation, accepté pour publication dans American Journal of Systems and Software, Special Issue on Multidisciplinary Perspectives of Agent-based Systems

4) TODIRAŞCU, Amalia, ION, Radu, NAVLEA, Mirabela, LONGO, Laurence (2011) *French text preprocessing with TTL*, Proceedings of Romanian Academy – Series A (Mathematics, Physics, Technical Sciences, Information Science), ISSN : 1454-9069, Volume 12, Number 2 April - June 2011, pp. 151–158

5) LONGO, Laurence, TODIRASCU Amalia, (2011), *Une étude de corpus pour la détection automatique de thèmes*. Revue électronique Texte et corpus, Actes des 6èmes Journées de la Linguistique de Corpus (Lorient, sept. 2009).

Articles parus dans des revues sans comité de lecture :

Activités dans la composante

Colloques : Longo, Laurence, Todiraşcu, Amalia (2010) La saillance référentielle en discours pour la détection des thèmes, Colloque Saillance 2, Strasbourg, Novembre 2010

Journées d'études :

Séminaires : Séminaire FDT 2016 (1 présentation), Séminaire LiLPa 2014 (1 présentation)

Participations scientifiques

Communications avec actes :

1) Todirascu, Amalia, Francois, Thomas, Bernhard, Delphine, Gala, Nuria, Ligozat, Anne-Laure (2016). Are Cohesive Features Relevant for Text Readability Evaluation?, Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, December 2016, Osaka, Japan, pp. 987-997.
2) Todiraşcu, Amalia, Navlea, Mirabela (2015) Integrating Verb+Noun Collocations into a French - Romanian Lexical Alignment System for Law Domain, EUROPHRAS 2015, June 2015, Malaga, Spain. 2015
3) Todiraşcu, Amalia, Navlea, Mirabela (2015) Aligning Verb+Noun Collocations to Improve a French -Romanian FSMT System, MUMTTT workshop, Jul 2015, Malaga, Spain. 2015
4) Monti, Johanna, Todiraşcu, Amalia (2015) Multiword Units Translation Evaluation In Machine Translation: Another Pain In The Neck? MUMTTT workshop of EUROPHRAS conference, Jul 2015, Malaga, Spain. 2015
5) Todiraşcu, Amalia, Sanchez-Cardenas, Beatriz (2015) Caractériser les discours académiques et de vulgarisation : quelles propriétés ? Actes du TALN 2015, 22-25 juin 2015
6) Todiraşcu, Amalia, Bertrand, Guillaume (2014) ClassYN : classer les documents selon le genre textuel, Mejri Salah, Sfar Inès, Van Campenhoudt, Marc (eds), *Actes des Journées scientifiques du réseau LTT*, Paris, septembre 2011, Éditions des archives contemporaines, ISBN : 9782813001627, pp 520-530.
7) Todiraşcu, Amalia, François, Thomas, Gala, Nuria, Fairon, Cédrick, Ligozat, Anne-Laure, Bernhard, Delphine (2013) Coherence and Cohesion for the Assessment of Text Readability, Proceedings of 10th International Workshop on Natural Language Processing and Cognitive Science, Maseille, octobre 2013
8) Todiraşcu, Amalia, Sanchez-Cardenas, Beatriz (2013) Une analyse linguistique des genres juridiques pour la classification automatique, Actes du Colloque "Corpus et Outils en Linguistique, Langues et Parole", 3-5 juillet Strasbourg (http://corpus-lilpa.sciencesconf.org/resource/page/id/6).

9) Longo, Laurence, Todiraşcu, Amalia (2013) Vers une modélisation des chaînes de référence dans des textes non-narratifs, Actes du Colloque "Corpus et Outils en Linguistique, Langues et Parole", 3-5 juillet Strasbourg (http://corpus-lilpa.sciencesconf.org/resource/page/id/6)
10) Navlea, M., Todiraşcu, A., 2012, Using Cognates to Improve Lexical Alignment Systems, in Petr Sojka, Aleš Horák, Ivan Kopeček, and Karel Pala (eds.), *Text, Speech and Dialogue* (15th International Conference, TSD 2012, Brno, Czech Republic, September 3-7, 2012. Proceedings), *Lecture Notes in Computer Science*, Volume 7499, pp. 370-377, Springer Berlin Heidelberg, ISBN: 978-3-642-32789-6 (Print) 978-3-642-32790-2 (Online)
11) Todirascu, Amalia, Pado, Sebastian, Krisch, Jennifer, Kisselew, Max, Heid, Ulrich (2012) French and German Corpora for Audience-based Text Type Classification, Calzolari, Nicoletta, Choukri, Khalid Declerck, Thierry, Doğan, Mehmet Uğur, Maegaard, Bente, Mariani, Joseph, Moreno, Asuncion, Odijk, Jan, Piperidis, Stelios (eds) Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), 23-25 May 2012, Istanbul, Turkey, European Language Resources Association (ELRA), ISBN 978-2-9517408-7-7, http://www.lrec-conf.org/proceedings/lrec2012/index.html
12) Navlea, M., Todiraşcu, A., 2011, « Using Cognates in a French - Romanian Lexical Alignment System: A Comparative Study », in Galia Angelova, Kalina Bontcheva, Ruslan Mitkov, and Nikolai Nikolov (eds.), in Proceedings of the 8th International Conference on Recent Advances in Natural Language Processing (RANLP 2011), pp. 247-253, Hissar, Bulgaria, September 2011. INCOMA Ltd., Shoumen, Bulgaria ISSN 1313-8502. (idexed by ACL Anthology and DLBP Trier)
13) Navlea, Mirabela, Todiraşcu, Amalia, (2011) « Repérage automatique des équivalences traductionnelles pour un système de traduction automatique statistique français – roumain », Colloque Métiers et technologies de la traduction : quelles convergences pour l'avenir ? (TRALOGY 2011), Auditorium du CNRS, Paris, France, 3-4 mars 2011.
14) NAVLEA Mirabela, TODIRASCU, Amalia (2011) Cognate Identification for a French - Romanian Lexical Alignment System: Empirical Study, EAMT 2011, 30th-31th May 2011.
15) Navlea, Mirabela, Todiraşcu, Amalia (2011), « Ressources linguistiques pour un outil de traduction automatique statistique factorisée français – roumain », eds. Van Campenhoudt, Marc, Lino, Teresa, Costa Rute, Actes du Colloque « Passeurs de mots, passeurs d'espoir: lexicologie, terminologie et traduction face au défi de la diversité », 8èmes Journées Scientifiques, Réseau LTT, Agence Universitaire de la Francophonie - AUF, Lisbonne, Portugal, 15-17 octobre 2009, pp 375-389.
16) Laurence Longo and Amalia Todirascu (2011), RefGen: Identifying Reference Chains to Detect Topics, Series Ed.: Kacprzyk, Janusz “Advances in Intelligent and Soft Computing”, Springer Verlag, ISSN: 1860-949X. 1) Longo, Laurence, Todirascu, Amalia (2010): RefGen: a Tool for Reference Chains Identification. IMCSIT 2010: 447-454 (indexé par DBLP-Trier)

Communications sans actes :

Conférences invitées :

1) Invitation à l’Université Marne-La-Vallée, présentation dans le cadre du séminaire de l’équipe LIGM Laboratoire d'informatique Gaspard-Monge (mars 2012)
2) Invitation à l'Université Aix*Marseille, présentation dans le cadre du séminaire de l'équipe TALEP du Laboratoire d'Informatique Fondamentale de Marseille (octobre 2013)

Participation à des jurys

Jury de thèse: Geir Solskinnsbakk (NTNU, Trondheim, novembre 2012)

Jury HDR :

ACTIVITES D’ENSEIGNEMENT

Etablissement : Université de Strasbourg

Discipline : Linguistique Informatique (Dictionnaires électroniques, Analyse de corpus, Linguistique initiation), Informatique (Traitement Automatique des Langues, Bases de données,

Création de Sites Web, Recherches sur le Web, Genie Logiciel, Programmation, Analyse de corpus, Support Multimédia, XML, HTML)
Nature (CM, TD, TP) et volume (nombre d'heures effectives) : 247 HETD (CMs et TDs)
Niveau (L, M, D, à l'exception de la direction des thèses) :L, M
Enseignement "Outils d'exploration de corpus" pour l'Ecole doctorale des Humanités" (2008-2009, 2009-2010, 2013-2014, 2014-2015, 2015-2016)

RESPONSABILITES ADMINISTRATIVES

- Responsable pédagogique du Master Linguistique Informatique (2005-2009) et du Master Linguistique, Informatique, Traduction (depuis 2009)
- Directrice du Département informatique, UFR LSHA (octobre 2011 – août 2016)
- Responsable scientifique du projet en collaboration avec l'entreprise Rebuz (depuis juin 2013) qui concerne la thèse d' Yuylia Korencuk (sous contrat CIFRE, financé par l'ANRT). L'objectif du projet est de développer une méthode d'enrichissement des ontologies multilingues (anglais, français et allemand) à partir de corpus parallèles et comparables. Le projet propose des méthodes d'extraction de termes innovantes qui exploitent la structure des mots (les morphèmes, préfixes et suffixes), à l'aide des méthodes d'apprentissage automatique et de la traduction automatique. L'ontologie va être complétée par des relations entre termes extraits à partir des corpus comparables et parallèles.
- Responsable scientifique du projet CAP (collaboration avec le projet européen CLARIN - Common Language Ressource Infrastructure, avril 2009 – décembre 2010) vise à décrire les propriétés lexico-sémantiques de la fonction lexicale CAP (Mel'čuk (1984, 1988, 1992, 1999)). Cette fonction est un cas particulier de la relation d'accessibilité entre deux entités (personne, organisation). Nous utilisons ces résultats dans des applications d'extraction d'information. La description de la fonction, ainsi que les patrons d'extraction est définie sur la base des données linguistiques recueillies à partir de corpus multilingues (anglais, français, allemand).
- Responsable scientifique du projet CLASSYN (janvier 2010-décembre 2011, financé par le Ministère des Affaires Etrangères, en partenariat avec Université de Stuttgart et de Heidelberg). Ce projet a comme objectif d'identifier des propriétés syntaxiques, spécifiques au genre textuel, qui permettraient la classification automatique de documents (textes juridiques et débats parlamentaires, textes scientifiques et de vulgarisation), dans un contexte multilingue (français-allemand-anglais).
- Responsable scientifique du projet ANR ALECTOR (janvier 2017 – juin 2020). Ce projet vise à étudier les difficultés de lecture des enfants dyslexiques et de proposer un système d'aide à la lecture pour ces enfants ou personnes en difficulté, par le biais des textes simplifiés. Les ressources serviront aux enseignants et aux orthophonistes. Le projet regroupe le LIF, le LPL et le LPC (Université Aix*Marseille), le LIMSI (Paris Sud Orsay), LiLPa (Université de Strasbourg). Le projet regroupe spécialistes en linguistique, en didactique, en Traitement automatique des langues et en psychologie cognitive. Je suis responsable scientifique pour LiLPa, le LPL est le porteur du projet.

RESSOURCES ET OUTILS

A) corpus français étiqueté et lemmatisé corrigé manuellement (environ 900000 tokens), pour l'entraînement de l'étiqueteur TTL (Ion 2007) et de l'analyseur syntaxique de Bohnet (Bohnet 2009);
B) corpus parallèle aligné au niveau lexical FR-EN; EN-RO (1000 phrases);
C) RefGen – outil de détection automatique de chaînes de référence, développé en Java
(comprenant une base de patrons d'extraction des entités nommées et des emplois impersonnels de 'il');
D) Corpus annoté en relations de coréférence (multigenre : texte littéraire, rapport public, articles de presse) (20000 tokens), en format XML;
E) Corpus français de textes scientifiques-textes de vulgarisation (1000000 tokens), analysé syntaxiquement avec l'analyseur syntaxique de Bohnet (2009);
F) dictionnaire multilingue de collocations (français-roumain, 250 entrées);
G) prototype pour un dictionnaire bilingue pour la traduction français -espagnol (Transverb);
H) système de traduction factorisée français-roumain et roumain français