



# Gaël Harry Adélio André Dias

## Présentation

Prénoms et Nom	Gaël Harry Adélio André DIAS.
Date et Lieu de Naissance	2 Mars 1969 à Orléans (France).
Nationalité	Française.
Dernier Diplôme	Habilitation à Diriger des Recherches.
Qualification	Section 27.
Situation Professionnelle	Professeur des Universités (classe 1), Université de Caen Normandie, GREYC UMR 6072.
Mots Clefs	Traitement Automatique des Langues, Recherche d'Information, Interfaces Homme-Machine.

## Biographie

En 1993, j'obtiens une MIAGE à l'Université d'Orléans (UO) avec une moyenne de 15.07 (2° de promotion). Après mon service militaire, je prépare le DEA en Informatique (Langages, Programmation et Traduction) à l'UO et suis reçu avec une moyenne de 14.42 (6° de promotion) après la présentation de mon rapport de stage intitulé *An Approach to Machine Translation: An Interlingua Machine Translation System* sous l'orientation de Sylvie Billot. Je rejoins alors le groupe du Traitement Automatique des Langues (TAL) de l'Université de Sheffield (Angleterre) comme Assistant de Recherche pour une durée d'un an. En 1996, je reçois une invitation comme Assistant de Recherche au Centre d'Intelligence Artificielle de l'Université Nouvelle de Lisbonne (UNL, Portugal) dans lequel je reste jusqu'en 2002. En 1998, j'obtiens un bourse de doctorat (GGP XXI/BD/3895/96) financée par la Fondation pour la Science et la Technologie du Ministère Portugais de la Science, Technologie et Enseignement Supérieur. Après deux ans d'études à l'UNL, j'accepte un poste d'Attaché Temporaire d'Enseignement et de Recherche à l'Université de la Beira Interior (UBI, Portugal). En 2002, je suis reçu à l'unanimité du jury (classification maximale de l'UNL) au Doctorat en Informatique réalisé en co-tutelle avec l'UO et l'UNL sous la co-orientation de Jean-Claude Bassano de l'UO et de José Gabriel Pereira Lopes de l'UNL, dans lequel je présente mes travaux sur *L'Extraction Automatique d'Associations Lexicales à Partir de Corpora*. Je deviens alors Maître de Conférences au Département de Mathématique et Informatique à l'UBI et suis rattaché au Centre de Recherche en Mathématique de l'UBI. En 2003, je participe activement à la création du Département d'Informatique de l'UBI et je fonde en 2004 le Centre de Recherche en Technologie du Langage Humain et Bioinformatique. En 2008, je suis invité comme collaborateur à l'Institut de Recherche et Développement en Ingénierie des Systèmes et Logiciels de Lisbonne (INESC-ID, Portugal). En 2009, je participe à la création du cursus de 3° cycle d'études en Technologie du Langage Humain de l'Université de Porto (UP, Portugal) en tant que Maître de Conférences invité. En 2010, j'obtiens à l'unanimité du jury mon HDR à l'UO intitulée *Information Digestion* qui présente un nouveau paradigme de Recherche d'Information intégrant l'Acquisition de Connaissances Intrinsèques du Langage. Finalement, en 2011, je deviens Professeur des Universités au Département d'Informatique de l'Université de Caen Basse-Normandie (UCBN) et je prends la responsabilité du Master 2 en Traitement des Images et des Langues (IMALANG). En parallèle, j'intègre le Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen (GREYC CNRS UMR 6072) où je dirige l'équipe spécialisée en Technologie du Langage Humain (HULTECH).

Université de Caen Normandie – F-14032 Caen Cedex, France

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • 📠 +33 (0)2 31 56 73  
30 • ✉ [gael.dias@unicaen.fr](mailto:gael.dias@unicaen.fr) • 🌐 <https://dias.users.greyc.fr>

---

## Formation

Depuis mon DEA en Informatique (Langages, Programmation et Traduction) obtenu à l'UO, j'ai orienté toute ma formation autour du TAL et de la Recherche d'Information (RI). Mon rapport de stage de DEA propose une approche symbolique de traduction automatique *interlingua* basée sur une représentation sémantique compositionnelle typée. Ma thèse de Doctorat aborde le Traitement Statistique du Langage Naturel (TSLN) en proposant une méthode multilingue non supervisée d'extraction d'unités polylexicales. Finalement, mon HDR résume un nombre important de travaux de recherche autour du traitement de l'information explicite présente dans les textes et de l'acquisition de connaissances implicites du langage comme les ontologies.

### Habilitation à Diriger des Recherches (section 27)

- Titre *Information Digestion*.
- Lieu et Date Université d'Orléans. 2010.
- Jury Brigitte Grau (ENSIIE, Rapporteur), Marie-Francine Moens (Université Catholique de Louvain, Rapporteur), Manuel Vilares Ferro (Université de Vigo, Rapporteur), Adeline Nazarenko (Université Paris-Nord), Isabelle Tellier (UO), Mohand Boughanem (Université Toulouse III), Pierre Zweigenbaum (CNRS), Robert Mahl (Ecole des Mines de Paris).
- Mention A l'unanimité des membres du Jury.

### Doctorat en Informatique (co-tutelle)

- Titre Extraction Automatique d'Associations Lexicales à Partir de Corpora.
- Lieu et Date Université d'Orléans et Université Nouvelle de Lisbonne (Portugal). 1998-2002.
- Jury Pierre Zweigenbaum (CNRS, Rapporteur), António Porto (UNL, Président), Jean-Claude Basano (UO, Directeur), José Gabriel Pereira Lopes (UNL, Directeur), Teresa Lino (UNL), Christel Vrain (UO), Sylvie Billot (UO).
- Mention A l'unanimité des membres du Jury (classification maximale de l'UNL).

### DEA en Informatique

- Titre *An Approach to Machine Translation: An Interlingua Machine Translation System*.
- Mention Bien (6° de promotion). Université d'Orléans. 1995. Responsable: Sylvie Billot.

### Maîtrise en Informatique Appliquée à la Gestion (MIAGE)

- Mention Bien (2° de promotion). Université d'Orléans. 1993.

---

## Enseignement

Selon la classification ACM de 1998, j'enseigne (et ai enseigné) dans les domaines des Systèmes d'Information (section H) et des Méthodes Computationnelles (section I) au niveau du Master 1 et 2 en Traitement des Images et des Langues (IMALANG) de l'UCBN, du Diplôme d'Ingénieur de l'École Nationale Supérieure d'Ingénieurs de Caen (ENSICAEN), du Diplôme d'Ingénieur de l'École Supérieure d'Ingénieurs en Informatique et Génie des Télécommunications (ESIGETEL) et du Master 2 Sciences du Langage de l'UCBN. Je suis responsable depuis 2011 du Master 2 IMALANG commun à l'ENSICAEN.

## Section H: Systèmes d'Information

### Moteurs de Recherche

Classification	ACM (H.3.3).
Audience	Master 1 IMALANG de l'UCBN.
Heures	2h de CM et 2h de TP (40h/semestre).
Etudiants	10 étudiants/semestre en moyenne.
Responsabilités	Responsable et créateur de l'UE.
Années	2011-Présent.

### Recherche d'Information Multimédia

Classification	ACM (H.3.3).
Audience	Master 2 IMALANG de l'UCBN et Majeure Image et Multimédia de l'ENSICAEN.
Heures	2h de CM et 1h de TP (30h/semestre).
Etudiants	15 étudiants/semestre en moyenne.
Responsabilités	Chargé de Cours.
Années	2011-2013.

### Interface Homme-Machine

Classification	ACM (H.1.2).
Audience	Master 1 en Informatique de l'UCBN.
Heures	2h de CM et 2h de TP (40h/semestre).
Etudiants	30 étudiants/semestre en moyenne.
Responsabilités	Chargé de Cours.
Années	2012-2013.

## Section I: Méthodes Computationnelles

### Ingénierie des Langues

Classification	ACM (I.2.7).
Audience	Master 1 IMALANG de l'UCBN.
Heures	2h de CM et 2h de TP (40h/semestre).
Etudiants	15 étudiants/semestre en moyenne.
Responsabilités	Responsable et créateur de l'UE.
Années	2011-Présent.

### Ingénierie des Textes

Classification	ACM (I.2.7).
Audience	Master 1 IMALANG de l'UCBN.
Heures	2h de CM et 2h de TP (40h/semestre).
Etudiants	10 étudiants/semestre en moyenne.
Responsabilités	Responsable et créateur de l'UE.
Années	2011-Présent.

### Ingénierie des Connaissances Textuelles

Classification ACM (I.2.6).  
Audience Master 2 IMALANG de l'UCBN.  
Heures 2h de CM et 1h de TP (30h/semestre).  
Etudiants 15 étudiants/semestre en moyenne.  
Responsabilités Responsable et créateur de l'UE.  
Années 2011-Présent.

### Traitement Automatique des Langues

Classification ACM (I.2.7).  
Audience Master 2 en Sciences du Langage de l'UCBN.  
Heures 2h de CM (15h/semestre).  
Etudiants 5 étudiants/semestre en moyenne.  
Responsabilités Responsable et créateur de l'UE.  
Années 2012-Présent.

### Introduction au Traitement Automatique des Langues

Classification ACM (I.2.7).  
Audience Licence en Informatique de l'UCBN.  
Heures 2h de CM et 2h de TP (30h/semestre).  
Etudiants 25 étudiants/semestre en moyenne.  
Responsabilités Responsable et créateur de l'UE.  
Années 2012-Présent.

### Apprentissage Automatique

Classification ACM (I.2.6).  
Audience Master 2 IMALANG de l'UCBN et Majeure Image et Multimédia de l'ENSICAEN.  
Heures 2h de CM et 1h de TP (30h/semestre).  
Etudiants 15 étudiants/semestre en moyenne.  
Responsabilités Chargé de Cours.  
Années 2012-2013.

### Fouille de Données et Apprentissage Statistique

Classification ACM (I.2.6).  
Audience Master 1 en Informatique de l'UCBN.  
Heures 2h de CM et 2h de TP (40h/semestre).  
Etudiants 30 étudiants/semestre en moyenne.  
Responsabilités Chargé de Cours.  
Années 2012-2013.

### Exploration et Entrepôts de Données

Classification ACM (I.2.6).  
Audience Diplôme d'Ingénieur de l'ESIGETEL.  
Heures 2h de CM et 2h de TP (28h/semestre).  
Etudiants 10 étudiants/semestre en moyenne.

*Université de Caen Normandie – F-14032 Caen Cedex, France*

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • 📠 +33 (0)2 31 56 73  
30 • ✉ [gael.dias@unicaen.fr](mailto:gael.dias@unicaen.fr) • 🌐 <https://dias.users.greyc.fr>

Responsabilités Chargé de Cours.  
Années 2006-2012.

---

## Recherche

Mes recherches portent sur le TAL, l'Extraction de Connaissances et la Recherche d'Information contraintes par l'utilisation minimale de ressources linguistiques afin de proposer des solutions indépendantes de la langue. En particulier, j'étudie et j'encadre des travaux en Extraction d'Unités Polylexicales, Sémantique Lexicale, Construction d'Ontologies, Analyse des Sentiments, Recherche d'Information pour Dispositifs Nomades et Résumé de Textes. L'Extraction d'Unités Polylexicales est certainement l'axe de recherche que j'ai le plus développé en proposant en particulier un système hybride totalement indépendant de la langue et de valeurs seuils. Dans le cadre de la Sémantique Lexicale, j'ai proposé un nouveau paradigme de mesures de similarité informatives symétriques et asymétriques. En ce qui concerne la Construction d'Ontologies, j'ai travaillé sur la détection du niveau de généralité des termes à partir de l'application d'algorithmes de tri sur graphes orientés ainsi que sur le formalisme de prétopologie et étudié l'extraction automatique de synonymes à partir du repérage de paraphrases et de l'application de mesures de similarité distributionnelles. L'acquisition de connaissances intrinsèques au langage a permis de proposer des systèmes de Recherche d'Information pour Dispositifs Nomades, de Résumé de Textes et de Recherche d'Information Intelligente. Tout le travail théorique est intégré à un méta-moteur de recherche dans sa version pour dispositifs nomades.

## Publications

Je suis auteur et co-auteur de 129 publications en livres, chapitres de livres, revues dont 3 de classe A+/A (CS, IPM, JNLE), 1 de classe B (IR), conférences internationales dont 21 de classe A+/A (AAAI, ACL, CIKM, COLING, CoNLL, EACL, ECAI, ECML/PKDD, ESWC, KDD, MM, SIGIR, WWW), 14 de classe B (CICLING, EAMT, ECIR, EKAW, EPIA, IJCNLP, MT, SMC, SPIRE) et 18 de classe C (AWIC, EGC, FLAIRS, JADT, LREC, RANLP, TALN, WIC), et différents ateliers ou compétitions internationales (NTCIR, SEMEVAL) selon le système de classification CORE (*Computing Research and Education Association of Australasia*). La liste des abréviations se trouve en annexe I ainsi que la liste complète de toutes les publications dans la section Bibliographie.

## Orientations

J'ai encadré 6 thèses de Doctorat et suis actuellement co-directeur d'une thèse de Doctorat.

### Doctorat en Informatique de João Paulo Cordeiro

Titre *Rule Induction for Sentence Reduction*.  
Thème Une étude sur l'extraction non supervisée de règles de compression de phrases.  
Co-orientation Pavel Brazdil (UP, Portugal).  
Durée 2006-2011.  
Institution UBI, Portugal.  
Soutenance 29 Avril 2011.  
Jury Marie-Francine Moens (Université Catholique de Louvain, Belgique), Paulo Quaresma (Université d'Evora (UE), Portugal) et Maria Coheur (Université Technique de Lisbonne (UTL), Portugal).

### Doctorat en Informatique de Dinko Lambov

Titre *Cross Domain Multi-view Sentiment Classification*.  
Thème Un nouvel algorithme de classification multi-vues pour l'analyse de sentiments transversale aux domaines.  
Durée 2007-2011.  
Institution UBI, Portugal.  
Soutenance 22 Septembre 2011.

Université de Caen Normandie – F-14032 Caen Cedex, France

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • ☎ +33 (0)2 31 56 73  
30 • ✉ [gael.dias@unicaen.fr](mailto:gael.dias@unicaen.fr) • 🌐 <https://dias.users.greyc.fr>

Jury Mohand Boughanem (Université Paul Sabatier), Pavel Brazdil (UP, Portugal), Nuno Marques (UNL, Portugal), João Graça (Université de Pennsylvanie, EUA), Veska Noncheva (Université de Plovdiv, Bulgarie).

Financement Allocation Ministérielle Portugaise SFRH/BD/31532/2006.

#### Doctorat en Lettre d'Isabel Marcelino

Titre Lexique Médical Unifié pour le Portugais.

Thème Construction d'un dictionnaire électronique unifié du Portugais médical.

Orientation João Casteleiro (Université de Lisbonne (UL), Portugal).

Co-orientation Gaël Dias et José Martinez-de-Oliveira (UBI, Portugal).

Durée 2006-2012.

Institution UBI, Portugal.

Soutenance 29 Juin 2012.

Jury Maria da Graça Pinto (UP, Portugal), Teresa Lino (UNL, Portugal), António Santos Pereira (UBI, Portugal), Paulo Osório (UBI, Portugal), João Malaca Casteleiro (UL, Portugal) and José Martinez (UBI, Portugal).

Financement Allocation Ministérielle Portugaise SFRH/BD/29625/2006.

#### Doctorat en Informatique de Rumen Moraliyski

Titre *Discovery of Word Semantic Relations based on Sentential Context Analysis.*

Thème Une nouvelle méthodologie d'extraction de synonymes basée sur une analyse paradigmatique et syntagmatique.

Durée 2005-2013.

Institution UBI, Portugal.

Soutenance 22 Février 2013.

Jury Pablo Gamallo (Université de Santiago de Compostelle, Espagne), José Lopes (UNL, Portugal), Antoine Doucet (UCBN), António Branco (UL, Portugal).

Financement Allocation Ministérielle Portugaise SFRH/BD/19909/2004.

#### Doctorat en Informatique de Ricardo Campos

Titre *Disambiguating Implicit Temporal Queries for Temporal Information Retrieval Applications.*

Thème Une nouvelle mesure de désambiguïsation temporelle pour le recherche temporelle d'information.

Co-orientation Alípio Jorge (UP, Portugal).

Durée 2009-2013.

Institution UP, Portugal.

Soutenance 5 Juin 2013.

Jury Mário Silva (UTL, Portugal), Luís Torgo (UP, Portugal), Ricardo Baeza-Yates (Yahoo!), Sérgio Nunes (UP, Portugal).

Financement Allocation Ministérielle Portugaise SFRH/BD/63646/2009.

Prix Prix FRAUNHOFER Portugal de la meilleure thèse de Doctorat 2013.

#### Doctorat en Informatique de Sebastião Pais

Titre *Asymmetric Distributional Similarity Measures to Recognize Textual Entailment by Generality.*

Thème Une nouvelle méthode non supervisée d'inférence textuelle par généralité.

Co-orientation Robert Mahl (Ecole des Mines de Paris) et Kasia Weryn-Wolska (ESIGETEL).

Université de Caen Normandie – F-14032 Caen Cedex, France

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • 📠 +33 (0)2 31 56 73  
30 • ✉ [gael.dias@unicaen.fr](mailto:gael.dias@unicaen.fr) • 🌐 <https://dias.users.greyc.fr>

Durée 2009-2013.  
Institution UBI (Portugal) et Ecole des Mines de Paris.  
Soutenance 6 Décembre 2013.  
Jury Witold Kosinski (Polish-Japanese Institute of Information Technology, Poland), Antoine Doucet (UCBN), João Paulo Cordeiro (UBI, Portugal) et Zornitsa Kozareva (University of Southern California, USA).

#### Doctorat en Informatique de José Moreno Franco

Titre *Text-Based Ephemeral Clustering for Web Image Retrieval on Mobile Devices.*  
Thème Etude sur la classification non supervisée éphémère pour la recherche d'images de la toile pour dispositifs nomades.  
Durée 2011-2014.  
Institution UCBN.  
Soutenance 5 Juin 2013.  
Jury Adam Jatowt (University of Kyoto, Japan), Massih-Amini Reza (Université Joseph Fourier), Olivier Ferret (CEA), Guillaume Cleuziou (UO) et Béatrice Daille (Université de Nantes).  
Financement Allocation Ministérielle Française.

#### Doctorat en Informatique de Arthur Vaïsse-Lesteven

Titre *Ontological and Spatio-Temporal Reasoning for Augmented Data Mining.*  
Thème Introduction de raisonnement ontologique dans la fouille de données.  
Orientation Maroua Bouzid (UCBN).  
Durée 2013-2016.  
Institution UCBN et AIRBUS.  
Financement Allocation CIFRE.

## Projets Financés

J'ai été impliqué dans cinq projets nationaux (équivalents des projets ANR) financés par la Fondation pour la Science et la Technologie du Ministère Portugais de la Science, Technologie et Enseignement Supérieur (FCT). J'ai également mis en place trois projets bilatéraux financés par le programme Hubert Curien de l'EGIDE et la Conférence des Présidents d'Université. Depuis mon retour en France, je participe à 2 projets ANR et suis responsable d'un projet CPER.

#### ART-ADN - Accès par Retour Tactilo-oral Aux Documents Numériques

Source ANR.  
Appel *Innovative Societies : Innovation, Economy, Living.*  
Thème Développement d'une interface haptique pour l'accès aux documents du Web pour les populations malvoyantes.  
Durée 3 ans.  
Conclusion 2016.  
Responsable Fabrice Maurel (UCBN).  
Participants EA PALM 4649.

#### DOCSCOPE - Analyse Automatisée pour la Fraude Documentaire

Source ANR.  
Appel *Concepts, Systems and Tools for Global Security.*  
Thème Analyse théorique et sémiotique de la réalisation de faux dans les documents électroniques.

Université de Caen Normandie – F-14032 Caen Cedex, France

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • 📠 +33 (0)2 31 56 73  
30 • ✉ [gael.dias@unicaen.fr](mailto:gael.dias@unicaen.fr) • 🌐 <https://dias.users.greyc.fr>

Durée 3 ans.  
Conclusion 2015.  
Responsable Marc Pic (HI Hologram Industries Inc.).  
Participants CNRS GREYC UMR 6072, ExoMaquina Inc., Telecom Bretagne.

*SEMPCOMP - Utilisation de Ressources Sémantiques Componentielles dans des Situations d'Apprentissage Impliquant l'Utilisation de Documents*

Source Région Basse-Normandie.  
Référence 12P1512  
Thème Personnalisation de la Recherche d'Information à partir de la réalisation de modèles ontologiques d'utilisateurs.  
Durée 3 ans.  
Conclusion 2014.  
Responsable Gaël Dias et Stéphane Ferrari (GREYC).  
Participants Noopsis Inc..

*VIPACCESS - Ubiquitous Web Access for Visually Impaired People*

Source FCT.  
Référence PTDC/PLP/72142/2006.  
Thème Construction d'un méta-moteur de recherche pour dispositifs nomades qui intègre des fonctions permettant l'accès à l'information aux aveugles. En particulier, nous développons des techniques de Classification Ephémère, de Résumé de Textes et de Personnalisation sur le Web.  
Financement 126.055 euros.  
Durée 3 ans.  
Conclusion 2011.  
Responsable Gaël Dias (UBI, Portugal).  
Participants UP (Portugal), UNL (Portugal), *Massachusetts Institute of Technology* (EUA) et *University of North Texas* (EUA).

*PT-STAR - Speech Translation Advanced Research to and from Portuguese*

Source FCT.  
Référence CMU-PT/HuMach/0039/2008.  
Thème Construction d'un traducteur *Speech-to-Speech* entre le Portugais et l'Anglais. En particulier, j'ai participé au développement de techniques d'extraction automatique de thesauri bilingues à partir de l'identification de paraphrases et de leur alignement à partir d'algorithmes de bioinformatique.  
Financement 134.879 euros.  
Durée 3 ans.  
Conclusion 2011.  
Responsable Luisa Coheur (Université Technique de Lisbonne, Portugal).  
Participants Gaël Dias, UBI (Portugal) et *Carnegie Mellon University* (EUA).

*MEDON - Using Ontologies to Model Data and Medical Procedures*

Source FCT.  
Référence PTDC/EIA/80772/2006.

*Université de Caen Normandie – F-14032 Caen Cedex, France*

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • 📠 +33 (0)2 31 56 73 30 • ✉ [gael.dias@unicaen.fr](mailto:gael.dias@unicaen.fr) • 🌐 <https://dias.users.greyc.fr>



Thème Construction d'ontologies de procédures et données cliniques. En particulier, j'ai participé à la construction automatique d'ontologies à partir du formalisme de prétopologie et à la création du dictionnaire unifié du Portugais médical.

Financement 97.366 euros.

Durée 3 ans.

Conclusion 2011.

Responsable Paulo Quaresma (Université de Evora, Portugal).

Participants Gaël Dias, UBI (Portugal) et UNL (Portugal).

#### *SUMO - Automatic Text Summarization for Mobile Technologies*

Source FCT.

Référence POSC/PLP/57438/2004.

Thème Evaluer les capacités de techniques de Résumé de Textes pour la visualisation d'information sur dispositifs nomades.

Financement 57.098 euros.

Durée 3 ans.

Conclusion 2008.

Responsable Gaël Dias (UBI, Portugal).

Participants UP (Portugal).

#### *SITE-O-MATIC - Web Automation and Adaptive Web*

Source FCT.

Référence POSC/EIA/58367/2004.

Thème Adaptation automatique de sites web à partir d'analyses de logs. En particulier, j'ai participé dans le cadre de l'organisation de l'information textuelle par Classification Ephémère.

Financement 76.999 euros.

Durée 3 ans.

Conclusion 2008.

Responsable Alípio Jorge (Université de Porto, Portugal).

Participants Gaël Dias, UBI (Portugal) et Université du Minho (UM) (Portugal).

#### *LEILA II - Learning Lexical Associations towards Ontology Construction*

Source Conférence des Présidents d'Université.

Référence F-48/07.

Thème Construction d'ontologies à partir de l'évaluation du degré de similarité et de généralité entre termes.

Financement 2.000 euros.

Durée 2 ans.

Conclusion 2009.

Responsables Gaël Dias (UBI, Portugal) et Guillaume Cleuziou (UO).

#### *LEILA - Learning Lexical Associations*

Source Conférence des Présidents d'Université.

Référence F-20/05.

Thème Applications d'algorithmes génétiques pour l'extraction d'unités polylexicales.

Financement 1.000 euros.

Université de Caen Normandie – F-14032 Caen Cedex, France

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • ☎ +33 (0)2 31 56 73  
30 • ✉ [gael.dias@unicaen.fr](mailto:gael.dias@unicaen.fr) • 🌐 <https://dias.users.greyc.fr>

Durée 1 an.  
Conclusion 2006.  
Responsables Gaël Dias (UBI, Portugal) et Guillaume Cleuziou (UO).

### *MULTILEXI - Multilingual Term Extraction for Lexicographic Purposes*

Source Programme Hubert Curien de l'EGIDE.  
Thème Construction de terminologies bilingues en Anglais et Slovène.  
Financement 1.000 euros.  
Durée 1 an.  
Conclusion 2006.  
Responsables Gaël Dias (UBI, Portugal) et Spela Vintar (Université de Ljubljana, Slovénie).

## Comités d'Expertise

Grâce à l'expérience acquise à travers mes différentes recherches dans le domaine du TAL et de la RI, j'ai régulièrement été sollicité dans le cadre de comités de conseils, de comités de rédaction et de comités de relecture. En particulier, je fais partie du comité de conseil de la Conférence Portugaise en Intelligence Artificielle depuis 2005. J'ai été membre du comité de rédaction des 3 livres suivants: *Applied Natural Language Processing and Content Analysis: Identification, Investigation, and Resolution* de Philip McCarthy et Chutima Boonthum publié dans l'IGI Global, *Handbook of NLP* de Nitin Indurkha publié chez Taylor & Francis CRC-Press et *Building an Intelligent Web: Theory and Practice* de Pawan Lingras publié chez John and Bartlett Publishers. Je suis également membre du comité de lecture de 9 revues internationales dont 6 de classe A+/A (ASLP, CS, IPM, JCSL, JNLE, ML) et 3 de classe B (DKE, JLRE, TSLP) et de 15 conférences et workshops internationaux dont 6 de classe A+/A (ACL, COLING, EAACL, EMNLP, HLT-NAACL, SIGIR), 3 de classe B (ECIR, EPIA, IJCNLP) et 6 de classe C (AWIC, EGC, FLAIRS, ICAART, LREC, RANLP). La liste complète et détaillée se trouve ci-dessous (la liste des abréviations se trouve en annexe I).

ML 2013 [CORE = A+, Thomson Reuters Citation Index Impact Factor = 1.689].  
CS 2013 [CORE = A+, Thomson Reuters Citation Index Impact Factor = 4.043].  
IPM 2006-Present [CORE = A, Thomson Reuters Citation Index Impact Factor = 1.783].  
JCSL 2004 [CORE = A, Thomson Reuters Citation Index Impact Factor = 1.09].  
JNLE 2006 [CORE = A, Thomson Reuters Citation Index Impact Factor = 0.47].  
ASLP 2007 [CORE = A, Thomson Reuters Citation Index Impact Factor = 2.625].  
JLRE [2008 CORE = B, Thomson Reuters Citation Index Impact Factor = 0.634].  
TSLP 2012 [CORE = B, Thomson Reuters Citation Index Impact Factor = 2.625].  
DKE 2010, 2011, 2012, 2013 [CORE = B, Thomson Reuters Citation Index Impact Factor = 1.745].  
ACL 2004, 2006, 2007, 2009, 2011 [CORE = A+, Conference Impact Factor = 0.90].  
SIGIR 2005 [CORE = A+, Conference Impact Factor = 0.96].  
HLT-NAACL 2006, 2007, 2011, 2012, 2013 [CORE = A, Conference Impact Factor = 0.88].  
COLING 2006, 2010, 2012 [CORE = A, Conference Impact Factor = 0.64].  
ECIR 2006 [CORE = B, Conference Impact Factor = 0.64].  
EPIA 2005, 2007, 2009, 2011, 2013 [CORE = B].  
IJCNLP 2009 [CORE = B].  
RANLP [CORE = C, Conference Impact Factor = 0.54].  
AWIC 2009 [CORE = C].  
EGC 2011, 2010 [CORE = C].  
FLAIRS 2008, 2009, 2010 [CORE = C].  
ICAART 2011, 2010, 2009 [CORE = C].  
LREC 2008, 2014 [CORE = C].

Université de Caen Normandie – F-14032 Caen Cedex, France

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • ☎ +33 (0)2 31 56 73

30 • ✉ [gael.dias@unicaen.fr](mailto:gael.dias@unicaen.fr) • 🌐 <https://dias.users.greyc.fr>

## Jurys de Thèses de Doctorat

J'ai participé à de nombreux jurys de thèses de HDR et de Doctorat en informatique. En particulier, j'ai participé à 4 thèses de HDR et 20 thèses de doctorat tant au niveau national comme international.

*De la Modélisation Linguistique à la Recherche Documentaire.*

Etudiant Tasadit Amghar.  
Université Université d'Angers.  
Date 2012.  
Niveau HDR en Informatique (section 27).  
Responsabilité Rapporteur.

*Modélisation de l'Architecture Textuelle : Apport à (et de) la Présentation, l'Interaction, et l'Accessibilité en Situation Normale et Dégradée*

Etudiant Mustapha Mojahid.  
Université Université Paul Sabatier Toulouse III.  
Date 2013.  
Niveau HDR en Informatique (section 27).  
Responsabilité Examineur.

*A Framework for Temporal Web Analytics*

Etudiant Marc Spaniol.  
Université Université de Caen Basse-Normandie.  
Date 2014.  
Niveau HDR en Informatique (section 27).  
Responsabilité Parrain.

*Extraction, Exploitation and Evaluation of Document-based Knowledge*

Etudiant Antoine Doucet.  
Université Université de Caen Basse-Normandie.  
Date 2012.  
Niveau HDR en Informatique (section 27).  
Responsabilité Parrain.

*Construction d'une Cartographie de Domaine à partir de Ressources Sémantiques Hétérogènes*

Etudiant Sarra Ben Abbès.  
Maître de Thèse Adeline Nazarenko.  
Université Université Paris XIII.  
Date 2013.  
Niveau Doctorat en Informatique.  
Responsabilité Rapporteur.

*Processing Temporal Information in Unstructured Documents*

Etudiant Francisco Costa.

Maître de Thèse António Branco.  
Université Université de Lisbonne (Portugal).  
Date 2013.  
Niveau Doctorat en Informatique.  
Responsabilité Rapporteur.

*Unsupervised Information Extraction from Text: Extraction and Clustering of Relations between Entities*

Etudiant Wei Wang.  
Maître de Thèse Brigitte Grau, Olivier Ferret et Romaric Besançon.  
Université Université Paris Sud.  
Date 2013.  
Niveau Doctorat en Informatique.  
Responsabilité Rapporteur.

*Adquisición y Representación del Conocimiento Mediante Procesamiento del Lenguaje Natural*

Etudiant Milagros Fernandez Gavilanes.  
Maître de Thèse Manuel Vilares Ferro et Eric de la Clergerie.  
Université Université de la Corogne (Espagne).  
Date 2012.  
Niveau Doctorat en Informatique.  
Responsabilité Rapporteur.

*A Generic and Open Framework for Multiword Expressions Treatment: from Acquisition to Applications*

Etudiant Carlos Ramisch.  
Maître de Thèse Christian Boitet et Aline Villavicencio.  
Université Université de Grenoble.  
Date 2012.  
Niveau Doctorat en Informatique.  
Responsabilité Rapporteur.

*Term Weighting in Short Documents for Categorization and Keyword Extraction*

Etudiant Mika Timonen.  
Maître de Thèse Hannu Toivonen.  
Université Université de Helsinki (Finlande).  
Date 2012.  
Niveau Doctorat en Informatique.  
Responsabilité Rapporteur.

*Contextualização e Ativação em Contexto de Sistemas de Perguntas e Respostas*

Etudiant José Sais.  
Maître de Thèse Paulo Quaresma.  
Université Université d'Evora (Portugal).

Université de Caen Normandie – F-14032 Caen Cedex, France

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • 📠 +33 (0)2 31 56 73  
30 • ✉ gael.dias@unicaen.fr • 🌐 https://dias.users.greyc.fr

Date 2010.  
Niveau Doctorat en Informatique.  
Responsabilité Rapporteur.

*A Text Mining System for Evaluating the Stock Market's Response To News*

Etudiant Brett Drury.  
Maître de Thèse Luís Torgo et José João Dias de Almeida.  
Université Université de Porto (Portugal).  
Date 2010.  
Niveau Doctorat en Informatique.  
Responsabilité Rapporteur.

*Analyse des Sentiments: Système Autonome d'Exploration des Opinions Exprimées dans les Critiques Cinématographiques*

Etudiant Grzegorz Dzikowski.  
Maître de Thèse Robert Mahl.  
Université Ecole des Mines de Paris.  
Date 2008.  
Niveau Doctorat en Informatique.  
Responsabilité Rapporteur.

*Advanced Document Description, a Sequential Approach*

Etudiant Antoine Doucet.  
Maître de Thèse Helena Ahonen-Myka.  
Université Université d'Helsinki (Finlande).  
Date 2005.  
Niveau Doctorat en Informatique.  
Responsabilité Rapporteur.

*Uma Interface Sintaxe/Semântica Baseada em Regras Hierarquicamente Organizadas*

Etudiante Maria Coheur.  
Maître de Thèse Nuno Mamede.  
Université Université Technique de Lisbonne (Portugal).  
Date 2004.  
Niveau Doctorat en Informatique.  
Responsabilité Rapporteur.

*Evolutionary Approach for Entity Extraction and Anaphora Resolution*

Etudiant Utpal Kumar Sikdar.  
Maître de Thèse Asif Ekbal.  
Université Indian Institute of Technology - Patna.  
Date 2016.  
Niveau Doctorat en Informatique.  
Responsabilité Membre du Jury.

Vers une Capitalisation des Connaissances orientée Utilisateur: Extraction et Structuration Automatiques de l'Information Issue de Sources Ouvertes

Etudiant Laurie Serrano.  
Maître de Thèse Maroua Bouzid.  
Université Université de Caen Basse-Normandie.  
Date 2014.  
Niveau Doctorat en Informatique.  
Responsabilité Membre du Jury.

Veille Épidémiologique Multilingue : une Approche Parcimonieuse au Grain Caractère Fondée sur le Genre Textuel

Etudiant Gaël Lejeune.  
Maître de Thèse Nadine Lucas et Antoine Doucet.  
Université Université de Caen Basse-Normandie.  
Date 2013.  
Niveau Doctorat en Informatique.  
Responsabilité Membre du Jury.

Analyse Automatique d'Opinion : Problématique de l'Intensité et de la Négation pour l'Application à un Corpus Journalistique

Etudiant Lei Zhang.  
Maître de Thèse Stéphane Ferrari et Patrice Enjalbert.  
Université Université de Caen Basse-Normandie.  
Date 2012.  
Niveau Doctorat en Informatique.  
Responsabilité Membre du Jury.

Conception et Validation d'une Méthode de Complétion des Valeurs Manquantes Fondée sur leurs Modèles d'Apparition

Etudiant Leila Ben Othman Amroussi.  
Maître de Thèse Bruno Crémilleux, François Rioult et Sadok Ben Yahiat.  
Université Université de Caen Basse-Normandie.  
Date 2011.  
Niveau Doctorat en Informatique.  
Responsabilité Membre du Jury.

Sydonie: Architecture de Document et Ingénierie du Web

Etudiant Jean-Marc Lecarpentier.  
Maître de Thèse Hervé Le Crosnier.  
Université Université de Caen Basse-Normandie.  
Date 2011.  
Niveau Doctorat en Informatique.  
Responsabilité Membre du Jury.

*Corpus Lingüísticos Estructurados de Grandes Dimensiones: Metodología e Sistemas de Recuperación de Información*

Etudiant Mario Rodríguez.  
Maître de Thèse Jorge Graña Gil.  
Université Université de la Corogne (Espagne).  
Date 2010.  
Niveau Doctorat en Informatique.  
Responsabilité Membre du Jury.

*AuTema-Dis: uma Arquitectura Computacional para Identificação da Temática Discursiva em Textos em Língua Portuguesa*

Etudiante Ana Leal.  
Maître de Thèse Paulo Quaresma.  
Université Université d'Evora (Portugal).  
Date 2009.  
Niveau Doctorat en Informatique.  
Responsabilité Membre du Jury.

*Integração Dinâmica de Artefactos em Sistemas de Diálogo*

Etudiant Porfírio Filipe.  
Maître de Thèse Nuno Mamede.  
Université Université Technique de Lisbonne (Portugal).  
Date 2007.  
Niveau Doctorat en Informatique.  
Responsabilité Membre du Jury.

## Communications

Afin de diffuser mes travaux de recherche, j'ai présenté 18 communications orales dans plusieurs Universités et Instituts nationaux et internationaux. Je présente ci-dessous la liste des communications les plus importantes.

*Temporal Web Information Retrieval: Text and Image*

Institution Université de Kyoto (Japon).  
Invitation Adam Jatowt.  
Date 2013.

*Temporal Information Retrieval*

Institution Université Paul Sabatier.  
Invitation Mustapha Mojahid.  
Date 2013.

*Information Retrieval: A Perspective for Ambient Intelligence*

Institution Université de Caen Basse-Normandie.  
Invitation Frédérick Lemarchand.  
Date 2013.

Université de Caen Normandie – F-14032 Caen Cedex, France

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • 📠 +33 (0)2 31 56 73  
30 • ✉ [gael.dias@unicaen.fr](mailto:gael.dias@unicaen.fr) • 🌐 <https://dias.users.greyc.fr>

*First Order Similarity Measures for the Extraction of Multiword Units*

Institution Université de Kyoto (Japon).  
Invitation Adam Jatowt.  
Date 2013.

*Temporal Disambiguation of Timely Implicit Queries*

Institution International Workshop on Big Data Analytics for the Temporal Web (Paris).  
Invitation Marc Spagniol.  
Date 2012.

*Collocation Extraction*

Institution Université de la Beira Interior (Portugal).  
Invitation Reina Pereira.  
Date 2012.

*Transverse Subjectivity Learning*

Institution Université de Kyoto (Japon).  
Invitation Adam Jatowt.  
Date 2012.

*Textual Similarity*

Institution Université de Kyoto (Japon).  
Invitation Adam Jatowt.  
Date 2012.

*Cross Domain Subjectivity Classification using Multi-View Learning*

Institution Information Science Institute de l'Université de Californie du Sud (USA).  
Invitation Zornitsa Kozareva.  
Date 2011.

*Information Digestion*

Institution Université de Caen Basse-Normandie.  
Invitation Antoine Doucet.  
Date 2011.

*Learning Natural Language*

Institution Académie des Sciences (Bulgarie).  
Invitation Galia Angelova.  
Date 2009.

*Word Ordering*

Institution Université d'Orléans.  
Invitation Guillaume Cleuziou.  
Date 2009.

Université de Caen Normandie – F-14032 Caen Cedex, France

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • 📠 +33 (0)2 31 56 73  
30 • ✉ [gael.dias@unicaen.fr](mailto:gael.dias@unicaen.fr) • 🌐 <https://dias.users.greyc.fr>



*Biology Based Alignments of Paraphrases for Sentence Compression*

Institution INESC-ID (Portugal).  
Invitation Isabel Trancoso.  
Date 2007.

*Web Intelligent Information Systems based on Web Mining Search Results*

Event *II Workshop on Statistics, Mathematics and Computation (Lisbonne, Portugal)*.  
Date 2006.

*Efficient Extraction of Multiword Units*

Institution Université d'Helsinki (Finlande).  
Invitation Helena Ahonen-Myka.  
Date 2005.

*Informative Topic Segmentation*

Event XI International Summer Conference on Probability and Statistics, Seminar of Statistical Data Analysis, Workshop on Industrial Statistics (Sozopol, Bulgarie).  
Date 19 Juin 2004.

*Searching the Web by Discovering and Clustering Related Terms*

Institution Institut Jozef Stefan (Slovénie).  
Invitation Marko Grobelnik.  
Date 2004.

*Extraction of Textual Units from Unrestricted Texts*

Event Advanced School on Artificial Intelligence (Lisbonne, Portugal).  
Date 2000.

## Diffusion

J'ai co-organisé 2 conférences nationales ainsi que 8 Workshops dans des conférences nationales et internationales. J'ai également été co-éditeur de l'édition spéciale de la revue *Traitement Automatique des Langues* intitulée *Passage à l'Echelle - Complexité, Algorithmique et Architecture*. Ce numéro traite des problèmes calculatoires rencontrés par l'explosion des données à traiter par les systèmes du TAL avec l'avènement de la Toile. La liste des événements organisés se trouve ci-dessous.

*22ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2015)*

Lieu et Date Caen, 22 au 25 Juin 2015.  
Co-organisateurs Gaël Dias et Nadine Lucas (UCBN).

*2nd International Workshop on Histoinformatics (HISTOINFORMATICS 2014)*

Conférence *6th International Conference on Social Informatics (SOCINFO 2014)*.  
Lieu et Date Barcelone (Espagne), 10 au 13 Novembre.

Université de Caen Normandie – F-14032 Caen Cedex, France

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • 📠 +33 (0)2 31 56 73  
30 • ✉ [gael.dias@unicaen.fr](mailto:gael.dias@unicaen.fr) • 🌐 <https://dias.users.greyc.fr>

Co-organisateurs Gaël Dias, Adam Jatowt (Université de Kyoto, Japon), Marten Düring (Centre Virtuel de la Connaissance sur l'Europe, Luxembourg) et Antal van Den Bosch (Radboud University Nijmegen, Hollande).

*1st International Workshop on Histoinformatics (HISTOINFORMATICS 2013)*

Conférence *5th International Conference on Social Informatics (SOCINFO 2013)*.

Lieu et Date Kyoto (Japon), 25 au 27 Novembre.

Co-organisateurs Gaël Dias, Adam Jatowt (Université de Kyoto, Japon), Agostini-Ouafi Viviana (UCBN), Christian Gudehus (Université de Flensburg, Allemagne), Günter Mühlberger (Université de Innsbruck, Autriche).

*5th Workshop on Text Mining and Applications (TEMA 2013)*

Conférence *16th Portuguese Conference on Artificial Intelligence (EPIA 2013)*.

Lieu et Date Angra do Heroísmo (Portugal), 9 au 12 Septembre.

Co-organisateurs Gaël Dias, Vítor Rocio (*Universidade Aberta*, Portugal), Joaquim Silva (UNL, Portugal).

*4th Workshop on Text Mining and Applications (TEMA 2011)*

Conférence *15th Portuguese Conference on Artificial Intelligence (EPIA 2011)*.

Lieu et Date Lisbonne (Portugal), 10 au 13 Octobre 2011.

Co-organisateurs Gaël Dias, Vítor Rocio (*Universidade Aberta*, Portugal), Joaquim Silva (UNL, Portugal).

*3rd Workshop on Text Mining and Applications (TEMA 2009)*

Conférence *14th Portuguese Conference on Artificial Intelligence (EPIA 2009)*.

Lieu et Date Aveiro (Portugal), 14 au 15 Octobre 2009.

Co-organisateurs Gaël Dias, José Gabriel Pereira Lopes (UNL, Portugal), Vítor Rocio (*Universidade Aberta*, Portugal), Joaquim Silva (UNL, Portugal).

*2nd Workshop on Text Mining and Applications (TEMA 2007)*

Conférence *13th Portuguese Conference on Artificial Intelligence (EPIA 2007)*.

Lieu et Date Guimarães (Portugal), 6 au 7 Décembre 2007.

Co-organisateurs Gaël Dias, José Gabriel Pereira Lopes (UNL, Portugal), Vítor Rocio (*Universidade Aberta*, Portugal), Joaquim Silva (UNL, Portugal).

*12th Portuguese Conference on Artificial Intelligence (EPIA 2005)*

Lieu et Date Covilhã (Portugal), 5 au 8 Décembre 2005.

Co-organisateurs Gaël Dias, Amílcar Cardoso (Université de Coimbra, Portugal) et Carlos Bento (Université de Coimbra, Portugal).

*Workshop on Methodologies and Evaluation of Lexical Cohesion Techniques in Real-world Applications - Beyond Bag of Words (ELECTRA 2005)*

Conférence *28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2005)*.

Lieu et Date Salvador da Bahia (Brésil), 19 Août 2005.

Université de Caen Normandie – F-14032 Caen Cedex, France

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • 📠 +33 (0)2 31 56 73  
30 • ✉ [gael.dias@unicaen.fr](mailto:gael.dias@unicaen.fr) • 🌐 <https://dias.users.greyc.fr>

Co-organisateurs Gaël Dias, Olga Vechtomova (Université de Waterloo, Canada) et Rosie Jones (Yahoo!, EUA).

Workshop on Methodologies and Evaluation of Multiword Units in Real-world Applications (MEMURA 2004)

Conférence *4th International Conference on Language Resources and Evaluation* (LREC 2004).

Lieu et Date Lisbonne (Portugal), 25 Mai 2004.

Co-organisateurs Gaël Dias, Spela Vintar (Université de Ljubljana, Slovénie) et José Gabriel Pereira Lopes (UNL, Portugal).

## Prix

Les recherches que j'ai menées et encadrées avec mes collègues et étudiants ont été récompensées à plusieurs reprises. La liste des prix est énoncée ci-dessous.

Prix de la meilleure thèse de doctorat au concours FRAUNHOFER Portugal Challenge 2013

Titre *Disambiguating Implicit Temporal Queries for Temporal Information Retrieval Applications.*

Etudiant Ricardo Campos.

Directeur de thèse Gaël Dias.

Prix du Meilleur Article Scientifique [50]

Titre *Language-Independent Informative Topic Segmentation.*

Auteurs Gaël Dias et Elsa Alves.

Conférence *9th International Symposium on Social Communication.*

Lieu et Date Santiago de Cuba (Cuba), 24 au 28 Janvier 2005.

Prix du Meilleur Projet

Projet Résumé de textes automatique pour dispositifs nomades.

Événement *Microsoft Imagine Cup.*

Lieu et Date Lisbonne (Portugal), 3 Mai 2006.

Etudiant Bruno Conde.

Tuteur Gaël Dias.

Prix du 2ème Meilleur Projet

Projet Méta-moteur de recherche éducatif pour dispositifs nomades.

Événement *Microsoft Imagine Cup.*

Lieu et Date Lisbonne (Portugal), 31 Mai 2007.

Etudiants Hugo Costa, Fernando Cunha, Sebastião Pais et Victor Gonçalves.

Tuteur Gaël Dias.

Université de Caen Normandie – F-14032 Caen Cedex, France

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • ☎ +33 (0)2 31 56 73  
30 • ✉ [gael.dias@unicaen.fr](mailto:gael.dias@unicaen.fr) • 🌐 <https://dias.users.greyc.fr>

---

## Responsabilités Collectives

En 2003, le Département de Mathématique et d'Informatique de l'UBI se divise donnant naissance au nouveau Département d'Informatique. Je suis alors responsable de la structuration de la Licence et de la Maîtrise en Ingénierie Informatique pour les domaines des Logiciels (section D), des Systèmes d'Information (section H) et des Méthodes Computationnelles (section I).

En 2004, je crée le Centre de Technologie du Langage Humain et Bioinformatique qui regroupe des chercheurs en Mathématique, Sciences Cognitives et Linguistique Computationnelle de l'UBI.

En 2005, je deviens Vice-Président du Département d'Informatique pour une durée de deux ans. Pendant cette période, je m'occupe principalement des relations extérieures du Département et de la mise en oeuvre du processus de Bologne. Je mets donc en place deux programmes Erasmus avec l'UO et l'ESIGETEL, et je prépare les deux premiers cycles en Ingénierie Informatique et en Technologie et Systèmes d'Information où ma contribution principale se centre sur les sections D, H et I.

En 2006, le processus de Bologne entre en vigueur et en 2007 je deviens responsable du nouveau Programme Doctoral en Informatique, poste que j'occupe jusqu'en 2009.

En 2011, alors que je rejoins le Département d'Informatique de l'Université de Caen Basse-Normandie, je prends la responsabilité du Master 2 en Traitement des Images et de la Langue (Master IMALANG). Parallèlement, je deviens chef de l'équipe HULTECH (Technologies du Langage Humain) au sein du CNRS GREYC UMR 6072.

En 2013, je coordonne l'axe transverse de recherche "Patrimoine" de la nouvelle fédération NORMASTIC (qui regroupe de LITIS de Rouen et le GREYC) à la frontière des sciences humaines et sociales, et sciences de l'information et des communications.

En 2014, je deviens membre du comité permanent de la conférence francophone sur le traitement automatique des langues (TALN-RECITAL) ainsi de membre du comité de rédaction de la revue TAL.

---

## Recommandations

Pour plus de renseignements sur mes activités, je ci-joints la liste suivante de Professeurs des Universités: Professeur Marie-Francine Moens (Université Catholique de Louvain, Belgique, sien.moens@cs.kuleuven.be), Professeur Pavel Brazdil (Université de Porto, Portugal, pbrazdil@liaad.up.pt), Professeur Pierre Zweigenbaum (CNRS, pz@limsi.fr), Professeur Adeline Nazarenko (Université Paris XIII, adeline.nazarenko@lipn.univ-paris13.fr) et Professeur Ruslan Mitkov (Université de Wolverhampton, Angleterre, r.mitkov@wlv.ac.uk).

---

## Bibliographie

### Ouvrages Edités

- [1] Jatowt A., Lim E., Ding Y., Miura A., Tezuka T., Dias G., Tanaka K., Flanagan A., Bing, and Dai T. *Proceedings of the 5th International Conference on Social Informatics (SOCINFO 2013)*. Springer LNCS, 2013. ISBN: 9783319032597.
- [2] I. Matveeva, G. Dias, and A. Hassan. *Proceedings of the 7th Workshop on Graph-based Methods for Natural Language Processing (ACL 2012)*. The Association for Computer Linguistics, 2012. ISBN: 978-1-937284-37-4. [CORE=A+].
- [3] Dias G., Nunes C., Cordeiro J.P., Moraliyski R., Marcelino I., Mukelov R., Campos R., Santos C., Alves E., Conde B., and Nonchev B. *Language Independent Methodologies to Tackle Multilinguality*, pages 33–41. Incoma Ltd. Shoumen, Bulgaria, in readings in multilinguality. selected papers from young researchers in bis-21++. galia angelova, kiril simov, milena slavcheva (editors) edition, 2006. ISBN: 954-91743-6-0.
- [4] Dias G. and Alves E. *Multilingual Topic Segmentation*, pages 229–253. Cambridge Scholars Press, linguistics in the twenty first century edition, 2006. ISBN: 1904303862.
- [5] Dias G. *Extracção Automática de Unidades Polilexicais para o Português*. Mercado de Letras, in a língua portuguesa no computador. collection as faces da lingüística aplicada. tony beber sardinha (eds) edition, 2005. ISBN 8575910442.

Université de Caen Normandie – F-14032 Caen Cedex, France

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • ☎ +33 (0)2 31 56 73

30 • ✉ gael.dias@unicaen.fr • 🌐 https://dias.users.greyc.fr

- [6] Dias G., Madeira S., and Lopes J.G.P. *Extracting Concepts from Dynamic Legislative Text Collections*, pages 5–17. Continuum International Publishing Group, in meaningful texts: the extraction of semantic information from monolingual and multilingual corpora. geoff barnbrook, pernilla danielsson and michaela mahlberg (eds) edition, 2005. ISBN 082647490X.
- [7] Dias G., Sousa S., and Crochemore M. *Scaling Natural Language Processing: Complexity, Algorithms and Architectures*, volume 46 of *Traitement Automatique des Langues*. Hermès Science Publications, 2005. ISBN: 978-2-7462-1527-6.
- [8] Bento C., Cardoso A., and Dias G. *Proceedings of the 12th Portuguese Conference on Artificial Intelligence*. Progress in Artificial Intelligence Serie. Springer LNAI 3008, 2005. ISSN: 0302-9743. [CORE=B].
- [9] Bento C., Cardoso A., and Dias G. *Proceedings of the 2005 Portuguese Conference on Artificial Intelligence*. IEEE Press, 2005. ISBN: 0780393651. [CORE=B].
- [10] Jones R., Vechtomova O., and Dias G. *Proceedings of the Workshop on Methodologies and Evaluation of Lexical Cohesion Techniques in Real-world Applications - Beyond Bag of Words (SIGIR 2005)*. ACM editions, 2005. ISBN: 1595930345. [CORE=A+].
- [11] Dias G., Lopes J.G.L., and S. Vintar. *Proceedings of the Workshop on Methodologies and Evaluation of Multiword Units in Real-world Applications (LREC 2004)*. ELRA editions, 2004. ISBN: 2951740816, EAN 0782951740815.

#### Revue d'Audience Internationale avec Comité de Rédaction

- [12] Campos R., Dias G., Jorge A., and Nunes C. Identifying top relevant dates for implicit time sensitive queries. *Information Retrieval Journal*, 2017. [CORE B, IF=0.90]. ISSN: 1573-7659.
- [13] Campos R., Dias G., Jorge A., and Nunes C. Gte-rank: a time-aware search engine to answer time-sensitive queries. *Information Processing and Management*, 2016. [CORE=A, IF=1.27]. ISSN: 0306-4573.
- [14] Campos R., Dias G., Jorge A., and Jatowt A. Survey of temporal information retrieval and related applications. *ACM Computing Surveys*, 47(2):1–41, 2014. [CORE=A+, IF=4.04]. ISSN: 0360-0300.
- [15] Lambov D., Pais S., and Dias G. Merged agreement algorithms for domain independent sentiment analysis. *Procedia Social and Behavioral Sciences. Selected Papers of the 12th Conference of the Pacific Association for Computational Linguistics*, pages 248–257, 2011. ISSN: 1877-0428.
- [16] Dias G., Pais S., Wegrzyn-Wolska K., Mahl R., and Jouvelot P. Textual entailment by generality. *Procedia Social and Behavioral Sciences. Selected Papers of the 12th Conference of the Pacific Association for Computational Linguistics*, pages 258–266, 2011. ISSN: 1877-0428.
- [17] Dias G., Moraliyski R., Cordeiro J.P., Doucet A., and Ahonen-Myka H. Automatic discovery of word semantic relations using paraphrase alignment and distributional lexical semantics analysis. *Journal of Natural Language Engineering (JNLE 2010)*, 16(4):439–467, 2010. [CORE=A].
- [18] Campos R., Dias G., Nunes C., and Nonchev B. Clustering of web page search results: A full text based approach. *International Journal of Computer and Information Science (IJCIS 2008)*, 9(4):1525–1539, 2008. ISSN: 1525-9293.
- [19] Cordeiro J.P., Dias G., and Brazdil P. New functions for unsupervised asymmetrical paraphrase detection. *Journal of Software (JSW 2007)*, pages 12–23, 2007. ISSN: 1796-217X.
- [20] Moralyiski R. and Dias G. Combination of global and local attributional similarities for synonym detection. *A Pliska Studia Mathematica Bulgarica Journal*, N. Yanev (eds), 18:239–254, 2007. ISSN: 02049805.
- [21] Dias G., Alves E., and Nunes C. Topic segmentation: How much can we do by counting words and sequences of words? *A Pliska Studia Mathematica Bulgarica Journal*, N. Yanev (eds), 17:39–70, 2005. ISSN: 02049805.

Université de Caen Normandie – F-14032 Caen Cedex, France

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • ☎ +33 (0)2 31 56 73

30 • ✉ gael.dias@unicaen.fr • 🌐 <https://dias.users.greyc.fr>

- [22] Dias G. and Kaalep H. Automatic extraction of multiword units for estonian: Phrasal verbs. *Languages in Development*, H. Metslang and M. Rannut (eds), *Linguistics Edition 41*, Lincom-Europa, München, Germany, pages 81–91, 2003. ISBN: 3895867039.
- [23] Dias G., Guilloire S., Bassano J.C., and Lopes J.G.P. Extraction automatique d'unités lexicales complexes: Un enjeu fondamental pour la recherche documentaire. *Traitement Automatique des Langues*, Christian Jacquemin (eds), Paris, France, 41(2):447–473, 2000. [CORE=C]. ISBN: 2746202255.

Communications d'Audience Internationale avec Comité de Sélection

- [24] G. Balikas, G. Dias, M-R. Amini, and H. Akhmouch. Multi-task learning for semantic relations discovery. In *Proceedings of the 21st International Conference on Knowledge Engineering and Knowledge Management (EKAW 2018)*, 2018. [CORE=B].
- [25] Hasanuzzaman M., Dias G., and Way A. Demographic word embeddings for racism detection on twitter. In *Proceedings of the 8th International Joint Conference on Natural Language Processing (IJCNLP 2017)*, 2017. [CORE=B].
- [26] Hasanuzzaman M., Sze W.L., Salim M.P., and Dias G. Collective future orientation and stock markets. In *Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI 2016)*, 2016. [CORE=A].
- [27] Hasanuzzaman M. and Ferrari S. and Dias G. Transductive learning for the identification of word sense temporal orientation. In *Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI 2016)*, 2016. [CORE=A].
- [28] Hasanuzzaman M., Dias G., Ferrari S., Mathet Y., and Way A. Identifying temporal orientation of word senses. In *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL 2016)*, 2016. [CORE=A].
- [29] Moreno J. and Dias G. Search intent mining by word vectors clustering at ntcir-imine. In *Proceedings of the iMine Subtask of the NII Testbeds and Community for Information Access Research Workshop (NTCIR 2016)*, 2016.
- [30] S. Acharya, A. Ekbal, J. Moreno, S. Sriparna, G. Dias, and P. Santhanam. Multi-objective optimization for word sense induction based on content and interlink connections. In *Proceedings of the 21st International Conference on Applications of Natural Language to Information Systems (NLDB 2016)*, 2016. [CORE=C].
- [31] Hasanuzzaman M., Dias G., Ferrari S., and Mathet Y. Iterative tempowordnet. In *Proceedings of the 13th Extended Semantic Web Conference (ESWC 2016)*, 2016. [CORE=A].
- [32] Lecarpentier J.M., Manishina E., Maurel F., Ferrari S., Giguët E., Dias G., and Busson M. Tag thunder: Web page skimming in non visual environment using concurrent speech. In *Proceedings of the 7th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT 2016) associated to INTERSPEECH 2016*, 2016. [CORE=A+].
- [33] Safi W., Maurel F., Routoure J-M., Beut P., and Dias G. An empirical study for examining the performance of visually impaired people in recognizing shapes through a vibro-tactile feedback. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2015)*, 2015.
- [34] Cleuziou G. and Dias G. Learning pretopological spaces for lexical taxonomy acquisition. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2015)*, 2015. [CORE=A].
- [35] Moreno J. and Dias G. Adapted b-cubed metrics to unbalanced datasets. In *Proceedings of the 38th Annual ACM SIGIR Conference (SIGIR 2015)*, 2015. [CORE=A+].
- [36] Hasanuzzaman M., Saha S., Dias G., and Ferrari S. Understanding temporal query intent. In *Proceedings of the 38th Annual ACM SIGIR Conference (SIGIR 2015)*, 2015. [CORE=A+].

- [37] Safi W., Maurel F., Routoure J-M., Beut P., and Dias G. Web-adapted supervised segmentation to improve a new tactile vision sensory substitution (tvss) technology. In *Proceedings of the 6th International Conference on Ambient Systems, Networks and Technologies (ANT 2015)*, 2015.
- [38] Safi W., Maurel F., Routoure J-M., Beut P., and Dias G. Supervised segmentation of web pages for vibro-tactile access on touch-screen devices. In *Proceedings of the 1st International Symposium on Web Algorithms (ISWAG 2015)*, 2015.
- [39] Cleuziou G., Buscaldi D., Dias G., Levorato V., and Largeron C. Qassit: A pretopological framework for the automatic construction of lexical taxonomies from raw texts. In *Proceedings of the International Workshop on Semantic Evaluation (SEMEVAL 2015)*, 2015.
- [40] M. Hasanuzzaman, G. Dias, and S. Ferrari. Hultech at the ntcir-11 temporalia task: Ensemble learning for temporal query intent classification. In *Proceedings of the Temporalia Subtask of the NII Testbeds and Community for Information Access Research Workshop (NTCIR-11 2014)*, 2014.
- [41] Dias G., Hasanuzzaman M., Ferrari S., and Mathet Y. Tempowordnet for sentence time tagging. In *Proceedings of the 4th ACM Temporal Web Analytics Workshop (TEMPWEB 2014) associated to 23rd International World Wide Web Conference (WWW 2014)*, pages 833–838, 2014. [CORE=A+].
- [42] Moreno G., Moraliyski R., Berrezoug A., and Dias G. Hultech: A general purpose system for cross-level semantic similarity based on anchor web counts. In *Proceedings of the International Workshop on Semantic Evaluation (SEMEVAL 2014) associated to 25th International Conference on Computational Linguistics (COLING 2014)*, 2014. [CORE=A].
- [43] Safi W., Maurel F., Routoure J-M., Beut P., and Dias G. Blind browsing on hand-held devices: Touching the web to understand it better. In *Proceedings of the Data Visualization Workshop (DataWiz 2014) associated to 25th ACM Conference on Hypertext and Social Media (HYPERTEXT 2014)*, 2014. [CORE=A].
- [44] Safi W., Maurel F., Routoure J-M., Beut P., and Dias G. A hybrid segmentation of web pages for vibro-tactile access on touch-screen devices. In *Proceedings of the 3rd Workshop on Vision and Language (VL 2014) associated to 25th International Conference on Computational Linguistics (COLING 2014)*, 2014. [CORE=A].
- [45] Moreno J. and Dias G. Hultech at the ntcir-11 imine task: Mining intents with continuous vectors models. In *Proceedings of the IMine Subtask of the NII Testbeds and Community for Information Access Research Workshop (NTCIR-11 2014)*, 2014.
- [46] Hasanuzzaman M., Dias G., Ferrari S., and Mathet Y. Propagation strategies for building temporal ontologies. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014)*, pages 6–11, 2014. [CORE=A].
- [47] Moreno J. and Dias G. Easy web search results clustering: When baselines can reach state-of-the-art algorithms. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014)*, pages 1–5, 2014. [CORE=A].
- [48] Campos R., Dias G., Jorge A., and Nunes C. Gte-cluster: A temporal search interface for implicit temporal queries. In *Proceedings of the 35th European Conference on Information Retrieval (ECIR 2014)*, pages 775–779, 2014. [CORE=B].
- [49] Moreno J., Dias G., and Cleuziou G. Query log driven web search results clustering. In *Proceedings of the 37th Annual ACM SIGIR Conference (SIGIR 2014)*, pages 777–786, 2014. [CORE=A+].
- [50] Moreno J. and Dias G. Pagerank-based word sense induction within web search results clustering. In *Proceedings of Digital Libraries (DL 2014)*, 2014. [CORE=A+].
- [51] Acharya S., Saha S., Moreno J., and Dias G. Multi-objective search results clustering. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014)*, 2014. [CORE=A].

Université de Caen Normandie – F-14032 Caen Cedex, France

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • ☎ +33 (0)2 31 56 73

30 • ✉ gael.dias@unicaen.fr • 🌐 <https://dias.users.greyc.fr>

- [52] Campos R., Dias G., Jorge A., and Nunes C. Gte-rank: Searching for implicit temporal query results. In *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM 2014)*, pages 1–41, 2014. [CORE=A].
- [53] Moreno J. and Dias G. Hultech at the ntcir-10 intent-2 task: Discovering user intents through search results clustering. In *Proceedings of the INTENT-2 Subtask of the NII Testbeds and Community for Information Access Research Workshop (NTCIR-10 2013)*, 2013.
- [54] Moreno J., Dias G., and Cleuziou G. Post-retrieval clustering using third-order similarity measures. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)*, pages 153–158, 2013. [CORE=A+].
- [55] Cordeiro J.P., Dias G., and Brazdil P. Rule induction for sentence reduction. In *Proceedings of the 16th Portuguese Conference in Artificial Intelligence (EPIA 2013)*, pages 528–539, 2013. [CORE=B].
- [56] Moreno J. and Dias G. Using text-based web image search results clustering to minimize mobile. In *Proceedings of the 34th European Conference on Information Retrieval (ECIR 2013)*, pages 775–779, 2012. [CORE=B].
- [57] Campos R., Jorge A., Dias G., and Nunes C. Disambiguating implicit temporal queries by clustering top relevant dates in web snippets. In *Proceedings of the 2012 IEEE/WIC/ACM International Conference on Web Intelligence (WIC 2012)*, pages 1–8, 2012. [CORE=C].
- [58] Campos R., Dias G., Jorge A., and Nunes C. Gte: A distributional second-order co-occurrence approach to improve the identification of top relevant dates in web snippets. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM 2012)*, pages 2035–2039, 2012. [CORE=A].
- [59] Dias G., Moreno J., Jatowt A., and Campos R. Temporal web image retrieval. In *Proceedings of the 19th edition of the International Symposium on String Processing and Information Retrieval (SPIRE 2012)*, pages 199–204, 2012. [CORE=B].
- [60] Maurel F., Dias G., Routoure J-M., Vautier M., Beust P., Molina M., and Sann C. Haptic perception of document structure for visually impaired people on handled devices. In *Proceedings of the 4th International Conference on Software Development for Enhancing Accessibility and Fighting Info-exclusion (DSAI 2012)*, pages 319–329, 2012.
- [61] Moreno J. and Dias G. Image results exploration using ephemeral clustering. In *Proceedings of the 2012 World Wide Web Conference (WWW 2012)*, 2012. [CORE=A+].
- [62] Buscaldi D., Cleuziou G., Dias G., and Levorato V. Exploitation de l’asymétrie entre termes pour l’extraction automatique de taxonomies à partir de textes. In *Proceedings of the 12th Conférence Internationale Francophone sur l’Extraction et la Gestion de Connaissance (EGC 2012)*, pages 345–356, 2012. [CORE=C].
- [63] Lambov D. and Dias G. Transverse subjectivity classification. In *Proceedings of the ACM Workshop on Issues of Sentiment Discovery and Opinion Mining of the 18th ACM Conference on Knowledge Discovery and Data Mining (KDD 2012)*, pages 1–8, 2012. [CORE=A].
- [64] Campos R., Dias G., Jorge A., and C. Nunes. Enriching temporal query understanding through date identification: How to tag implicit temporal queries? In *Proceedings of the 2nd Temporal Web Analytics Workshop of the 2012 World Wide Web Conference (WWW 2012)*, pages 41–48, 2012. [CORE=A].
- [65] Dias G., Pais S., Wegrzyn-Wolska K., and Mahl R. Textual entailment by generality. In *Proceedings of the 12th Conference of the Pacific Association for Computational Linguistics (PACLING 2011)*, 2011. [CORE=B].
- [66] Lambov D., Pais S., and Dias G. Merged agreement algorithms for domain independent sentiment analysis. In *Proceedings of the 12th Conference of the Pacific Association for Computational Linguistics (PACLING 2011)*, 2011. [CORE=B].



- [67] Dias G., Cleuziou G., and Machado D. Informative polythetic hierarchical ephemeral clustering. In *Proceedings of the 2011 IEEE/WIC/ACM International Conference on Web Intelligence (WIC 2011)*, pages 104–111, 2011. [CORE=C].
- [68] Dias G., Pais S., Wegrzyn-Wolska K., and Mahl R. Recognizing textual entailment by generality using informative asymmetric measures and multiword unit identification to summarize ephemeral clusters. In *Proceedings of the 2011 IEEE/WIC/ACM International Conference on Web Intelligence (WIC 2011)*, pages 284–287, 2011. [CORE=C].
- [69] Campos R., Dias G., and Jorge A. An exploratory study on how temporal information impact classification and clustering of future-related web documents. In *Proceedings of the 15th Portuguese Conference in Artificial Intelligence (EPIA 2011)*, pages 581–596, 2011. [CORE=B].
- [70] Drury B., Dias G., and Torgo L. Contextual classification strategy for polarity classification of direct quotations from financial news. In *Proceedings of the International Conference On Recent Advances in Natural Language Processing (RANLP 2011)*, pages 434–440, 2011. [CORE=C].
- [71] Cleuziou G., Buscaldi D., Levorato V., and Dias G. A pretopological framework for the automatic construction of semantic-lexical structures from texts. In *Proceedings of the 20th ACM Conference on Information and Knowledge Management (CIKM 2011)*, pages 2453–2456, 2011. [CORE=A].
- [72] Marcelino I., Dias G., Casteleiro J., and Martinez de Oliveira J. Semi-automatic construction of the unified medical lexicon for portuguese. In *Proceedings of the 3rd Congreso Internacional de Lingüística de Corpus (CILC 2011)*, 2011.
- [73] Moreno J. and Dias G. Using ephemeral clustering and query logs to organize web image search results on mobile devices. In *Proceedings of the International ACM Workshop on Interactive Multimedia on Mobile and Portable Devices of the ACM Multimedia Conference (ACMMM 2011)*, pages 33–38, 2011. [CORE=A].
- [74] Campos R., Jorge A., and Dias G. Using web snippets and web query-logs to measure implicit temporal intents in queries. In *Proceedings of the 2nd Workshop on Query Representation and Understanding of the 34th ACM Annual SIGIR Conference (SIGIR 2011)*, 2011. [CORE=A+].
- [75] Dias G., Campos R., and Jorge A. Future retrieval: What does the future talk about? In *Proceedings of the Workshop on Enriching Information Retrieval of the 34th ACM Annual SIGIR Conference (SIGIR 2011)*, 2011. [CORE=A+].
- [76] Campos R., Dias G., and Jorge A. What is the temporal value of web snippets? In *Proceedings of the 1st International Temporal Web Analytics Workshop of the International World Wide Web Conference (WWW 2011)*, pages 9–16, 2011. [CORE=A+].
- [77] Cleuziou G., Dias G., and Levorato V. Acquisition de structures lexico-sémantiques à partir de textes: un nouveau cadre de travail fondé sur une structuration prétopologique. In *Proceedings of the 11ème Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances (EGC 2011)*, pages 107–118, 2011. [CORE=C].
- [78] Grigonyté G., Cordeiro J.P., Moraliyski R., Dias G., and Brazdil P. A paraphrase alignment for synonym evidence discovery. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)*, pages 403–411, 2010. [CORE=A].
- [79] Lambov D., Dias G., and J.V. Graça. Multi-view learning for text subjectivity classification. In *Proceedings of the Workshop on Computational Approaches to Subjectivity and Sentiment Analysis of the 19th European Conference on Artificial Intelligence (ECAI 2010)*, 2010. [CORE=A].
- [80] Cleuziou G., Dias G., and V. Levorato. Modélisation prétopologique pour la structuration sémantico-lexicale. In *Proceedings of the 17èmes Rencontres de la Société Francophone de Classification (SFC 2010)*, pages 111–115, 2010.
- [81] Cordeiro J.P., Dias G., and Brazdil P. Unsupervised induction of sentence compression rules. In *Proceedings of the Workshop on Language Generation and Summarisation of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the*

*4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL/IJCNLP 2009)*, pages 15–22, 2009. [CORE=A+].

- [82] Lambov D., Dias G., and V. Noncheva. Sentiment classification across domains. In *Proceedings of the 14th Portuguese Conference in Artificial Intelligence (EPIA 2009)*, pages 622–633, 2009. [CORE=B].
- [83] Dias G. and Moraliyski R. Relieving polysemy problem for synonymy detection. In *Proceedings of the 14th Portuguese Conference in Artificial Intelligence (EPIA 2009)*, pages 610–621, 2009. [CORE=B].
- [84] Dias G., Pais S., Cunha F., Costa H., Machado D., Barbosa T., and Martins B. Hierarchical soft clustering and automatic text summarization for accessing the web on mobile devices for visually impaired people. In *Proceedings of the 22nd International FLAIRS Conference (FLAIRS 2009)*, pages 231–236, 2009. [CORE=C].
- [85] Campos R., Dias G., and A. Jorge. Disambiguating web search results by topic and temporal clustering: A proposal. In *Proceedings of International Conference on Knowledge Discovery and Information Retrieval (KDIR 2009)*, pages 292–296, 2009.
- [86] Machado D., Barbosa T., Pais S., Martins B., and Dias G. Universal mobile information retrieval. In *Proceedings of the 13th International Conference on Human Computer Interaction (HCI 2009)*, pages 345–354, 2009.
- [87] Dias G., Mukelov R., and Cleuziou G. Unsupervised graph-based discovery of general-specific noun relationships from web corpora frequency counts. In *12th International Conference on Natural Language Learning (CoNLL 2008)*, pages 97–104, 2008. [CORE=A].
- [88] Dias G., Mukelov R., and G. Cleuziou. Mapping general-specific noun relationships to wordnet hypernym/hyponym relations. In *Proceedings of the 16th International Conference on Knowledge Engineering and Knowledge Management Knowledge Patterns (EKAW 2008)*, pages 198–212, 2008. [CORE=B].
- [89] Dias G., Mukelov R., and Cleuziou G. Unsupervised learning of general-specific noun relations from the web. In *Proceedings of the 21th International FLAIRS Conference (FLAIRS 2008)*, pages 147–153, 2008. [CORE=C].
- [90] Moraliyski R. and Dias G. One sense per discourse for synonym detection. In *Proceedings of the International Conference On Recent Advances in Natural Language Processing (RANLP 2008)*, pages 383–387, 2008. [CORE=C].
- [91] Kozarava Z., Moraliyski R., and Dias G. Web people search with domain ranking. In *Proceedings of the 11th International Conference on Text, Speech and Dialogue (TSD 2008)*, pages 133–140, 2008.
- [92] Cleuziou G. and Dias G. Apprentissage de mesures de similarité sémantiques: Etude d'une variante de la mesure infosimba. In *Proceedings of the 1st Joint Meeting of the Société Francophone de Classification and the Classification and Data Analysis Group of the Italian Society of Statistics (SFC-CLADAG 2008)*, pages 233–236, 2008.
- [93] Marcelino I., Dias G., Casteleiro J., and Martinez J. Semi-controlled construction of the european portuguese unified medical language system. In *Proceedings of the Workshop on Finding the Hidden Knowledge: Text Mining for Biology and Medicine (FTHK 2008)*, 2008.
- [94] Dias G., Alves E., and Lopes J.G.P. Topic segmentation algorithms for text summarization and passage retrieval: An exhaustive evaluation. In *22nd Conference on Artificial Intelligence (AAAI 2007)*, pages 1334–1340, 2007. [CORE=A+].
- [95] Cordeiro J.P., Dias G., Cleuziou G., and Brazdil P. Biology based alignments of paraphrases for sentence compression. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrase of the 45th Annual Meeting of the Association of Computational Linguistics (ACL 2007)*, 2007. [CORE=A+].

Université de Caen Normandie – F-14032 Caen Cedex, France

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • ☎ +33 (0)2 31 56 73

30 • ✉ [gael.dias@unicaen.fr](mailto:gael.dias@unicaen.fr) • 🌐 <https://dias.users.greyc.fr>

- [96] Cordeiro J.P., Dias G., and Brazdil P. Unsupervised learning of paraphrase. In *Proceedings of the 8th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2007)*, 2007. [CORE=B].
- [97] Dias G. and Conde B. Accessing the web on handheld devices for visually impaired people. In *Proceedings of the 5th Atlantic Web Intelligence Conference (AWIC 2007)*, pages 80–87, 2007. [CORE=C].
- [98] Cordeiro J.P., Dias G., and Brazdil P. Learning paraphrases from wns corpora. In *Proceedings of the 20th International FLAIRS Conference (FLAIRS 2007)*, pages 193–198, 2007. [CORE=C].
- [99] Cordeiro J.P., Dias G., and Brazdil P. A metric for paraphrase detection. In *Proceedings of the 2nd International Multi-Conference on Computing in the Global Information Technology (ICCGI 2007)*, pages 7–13, 2007.
- [100] Dias G. and Conde B. Towards web browsing for visually impaired people. In *Proceedings of the 4th International Conference on Information Technology: New Generations (ITNG 2007)*, pages 493–496, 2007.
- [101] Dias G., Santos C., and Cleuziou G. Automatic knowledge representation using a graph-based algorithm for language-independent lexical chaining. In *Workshop on Information Extraction Beyond the Document of the Joint Conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics (COLING/ACL 2006)*, pages 36–47, 2006. [CORE=A+].
- [102] Dias G. and Conde B. Efficient text summarization for web browsing on mobile devices. In *Proceedings of the Workshop on Ubiquitous User Modeling of the 17th European Conference on Artificial Intelligence (ECAI 2006)*, pages 9–12, 2006. [CORE=A].
- [103] Campos R., Dias G., and Nunes C. Wise: Hierarchical soft clustering of web page search results based on web content mining techniques. In *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (WIC 2006)*, pages 301–305, 2006. [CORE=C].
- [104] Dias G. and Alves E. Unsupervised topic segmentation based on word co-occurrence and multiword units for text summarization. In *Proceedings of the ELECTRA Workshop of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2005)*, pages 41–48, 2005. [CORE=A+].
- [105] Campos R. and Dias G. Automatic hierarchical clustering of web pages. In *Proceedings of the ELECTRA Workshop of the 28th Annual International ACM SIGIR Conference (SIGIR 2005)*, pages 83–85, 2005. [CORE=A+].
- [106] Dias G. and Vintar S. Unsupervised learning of multiword units from part-of-speech tagged corpora: Does quantity means quality? In *Proceedings of the 12th Portuguese Conference on Artificial Intelligence (EPIA 2005)*, pages 669–680, 2005. [CORE=B].
- [107] Dias G. and Alves E. Discovering topic boundaries for text summarization based on word co-occurrence. In *Proceedings of the International Conference On Recent Advances in Natural Language Processing (RANLP 2005)*, pages 187–191, 2005. [CORE=C].
- [108] Dias G. and Alves E. Language-independent informative topic segmentation. In *Proceedings of the 9th International Symposium on Social Communication*, pages 588–592, 2005.
- [109] Dias G. and Nunes S. Evaluation of different similarity measures for the extraction of multiword units in a reinforcement learning environment. In *Proceedings of the Proceedings of the 4th International Conference On Languages Resources and Evaluation (LREC 2004)*, pages 1717–1721, 2004. [CORE=C].
- [110] Alexandre L., Pereira M., Madeira S., Cordeiro J.P., and Dias G. Web image indexing: Combining image analysis with text processing. In *Proceedings of the 5th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2004)*, 2004.

- [111] Pereira R., Crocker P., and Dias G. A parallel multikey quicksort algorithm for mining multiword unit. In *Proceedings of the Workshop on Methodologies and Evaluation of Multiword Units in Real-world Applications associated with the 4th International Conference On Languages Resources and Evaluation (LREC 2004)*, pages 17–24, 2004.
- [112] Dias G. Multiword unit hybrid extraction. In *Workshop on Multiword Expressions of the 41st Annual Meeting of the Association of Computational Linguistics (ACL 2003)*, pages 41–49, 2003. [CORE=A+].
- [113] Gil A. and Dias G. Using masks, suffix array-based data structures and multidimensional arrays to compute positional ngram statistics from corpora. In *Proceedings of the Workshop on Multiword Expressions of the 41st Annual Meeting of the Association of Computational Linguistics (ACL 2003)*, pages 25–33, 2003. [CORE=A+].
- [114] Gil A. and Dias G. Efficient mining of textual associations. In *Proceedings of the International Conference on Natural Language Processing and Knowledge Engineering (NLPKE 2003)*, pages 549–555, 2003.
- [115] Dias G., Kaalep H., and Muischnek K. Automatic extraction of verb phrases from annotated corpora: A linguistic evaluation for estonian. In *Proceedings of the Workshop on Collocation of the joint 39th Annual Meeting of the Association of Computational Linguistics and 10th Conference of the European Chapter of the Association of Computational Linguistics (EACL/ACL 2001)*, 2001. [CORE=A+].
- [116] Dias G. and Nunes S. Does natural selection apply to natural language processing? an experiment for multiword unit extraction. In *Proceedings of the 2001 IEEE Systems, Man, and Cybernetics Conference (IEEE SMC 2001)*, pages 205 – 212, 2001. [CORE=B].
- [117] Ribeiro A., Dias G., Lopes J.G.P., and Mexia J. Cognates alignment. In *Proceedings of the Machine Translation Summit VIII (MT 2001)*, pages 287–293, 2001. [CORE=B].
- [118] Dias G. and Nunes S. Combining evolutionary computing and similarity measures to extract collocations from unrestricted texts. In *Proceedings of the International Conference On Recent Advances in Natural Language Processing (RANLP 2001)*, pages 261–264, 2001. [CORE=C].
- [119] Dias G., Kaalep H., and Muischnek K. Automatic extraction of multiword units for estonian: a comparison between annotated and non-annotated corpora. In *Proceedings of the International Futuristic Conference on Language Development*, 2001.
- [120] Dias G., Guilloré S., and Lopes J.G.P. Mining textual associations in text corpora. In *Proceedings of the Workshop on Text Mining associated to the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2000)*, pages 92–95, 2000. [CORE=A].
- [121] Dias G., Guilloré S., and Lopes J.G.P. Extracting textual associations from part-of-speech tagged corpora. In *Proceedings of the European Association for Machine Translation Workshop on Harvesting Existing Resources (EAMT 2000)*, pages 4–11, 2000. [CORE=B].
- [122] Dias G., Guilloré S., and Lopes J.G.P. Extraction automatique d’associations textuelles à partir de corpora non traités. In *Proceedings of the 5èmes Journées Internationales d’Analyse Statistique des Données Textuelles (JADT 2000)*, pages 213–221, 2000. [CORE=C].
- [123] Dias G., Guilloré S., and Lopes J.G.P. Benefiting from multi-domain corpora for extracting terminologically relevant multiword lexical units. In *Proceedings of the 9th EURALEX International Congress (EURALEX 2000)*, pages 339–350, 2000.
- [124] Dias G., Madeira S., Guilloré S., and Lopes J.G.P. Extracting concepts from dynamic legislative text collections. In *Proceedings of the 5th TransEuropean Language Resource ans Infrastructure Seminar (TELRI 2000)*, 2000.
- [125] Dias G., Guilloré S., Bassano J.C., and Lopes J.G.P. Combining linguistics with statistics for multiword term extraction: A fruitful association? In *Proceedings of the 6ème Conférence sur la Recherche d’Informations Assistée par Ordinateur (RIAO 2000)*, pages 1–20, 2000.

- [126] Dias G., Guilloiré S., and Lopes J.G.P. Normalisation of association measures for multiword lexical unit extraction. In *Proceedings of the International Conference on Artificial and Computational Intelligence for Decision, Control and Automation in Engineering and Industrial Applications (ACIDCA 2000)*, pages 207–216, 2000.
- [127] Silva J., Dias G., Guilloiré S., and Lopes J.G.P. Using localmaxs algorithm for the extraction of contiguous and non-contiguous multiword lexical units. In *Proceedings of the 9th Portuguese Conference in Artificial Intelligence (EPIA 1999)*, pages 113–132, 1999. [CORE=B].
- [128] Dias G., Guilloiré S., and Lopes J.G.P. Language independent automatic acquisition of rigid multiword units from unrestricted text corpora. In *Proceedings of the 6ème Conférence Annuelle sur le Traitement Automatique des Langues Naturelles (TALN 1999)*, pages 333–339, 1999. [CORE=C].
- [129] Dias G., Guilloiré S., Vintar S., and Lopes J.G.P. Identifying and integrating terminologically relevant multiword units in the ijs-elan slovene-english parallel corpus. In *Proceedings of the 10th Computational Linguistics In the Netherlands (CLIN 1999)*, pages 29–40, 1999.
- [130] Dias G., Guilloiré S., and Lopes J.G.P. Mutual expectation: a measure for multiword lexical unit extraction. In *Proceedings of the VExTAL Venezia per il Trattamento Automatico delle Lingue (VEXTAL 1999)*, pages 133–139, 1999.
- [131] Dias G., Guilloiré S., and Lopes J.G.P. Multiword lexical units extraction. In *Proceedings of the International Conference on Machine Translation and Computer Language Information Processing*, pages 119–126, 1999.
- [132] Dias G., Guilloiré S., and Lopes J.G.P. Multilingual aspects of multiword lexical units. In *Proceedings of the Workshop on Language Technologies of the 32th Annual Meeting fo the Societas Linguistica Europea (SLE 1999)*, pages 11–21, 1999.

#### Thèses

- [133] Dias G. Information digestion, 2010. HDR Thesis. Université d'Orléans.
- [134] Dias G. *Extraction Automatique d'Associations Lexicales à Partir de Corpora*. PhD thesis, Université d'Orléans and Universidade Nova de Lisboa (Portugal), 2002.
- [135] Dias G. An approach to machine translation: An interlingua machine translation system. Master's thesis, Université d'Orléans, 1995.

---

## Annexe I: Abréviations

- AAAI *Association for the Advancement of Artificial Intelligence Conference.*
- ACL *Association for Computational Linguistics Conference.*
- AWIC *Atlantic Web Intelligence Conference.*
- ASLP *IEEE Transactions on Audio, Speech and Language Processing.*
- CICLING *Conference on Intelligent Text Processing and Computational Linguistics.*
- CIKM *International Conference on Information and Knowledge Management.*
- COLING *International Conference on Computational Linguistics.*
- CoNLL *Conference on Computational Natural Language Learning.*
- CS *ACM Computing Surveys.*
- DKE *Data and Knowledge Engineering Journal.*
- EACL *European Chapter of the Association for Computational Linguistics Conference.*
- EAMT *European Association for Machine Translation Conference.*
- ECAI *European Conference on Artificial Intelligence.*
- ECIR *European Conference on Information Retrieval.*
- EGC *Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances.*

*Université de Caen Normandie – F-14032 Caen Cedex, France*

☎ +33 (0)6 02 25 89 30 • ☎ +33 (0)2 31 56 74 87 • ☎ +33 (0)2 31 56 73

30 • ✉ [gael.dias@unicaen.fr](mailto:gael.dias@unicaen.fr) • 🌐 <https://dias.users.greyc.fr>

- EKAW *Conference on Knowledge Engineering and Knowledge Management.*
- EMNLP *Conference on Empirical Methods in Natural Language Processing.*
- EPIA *Portuguese Conference on Artificial Intelligence.*
- FLAIRS *International Florida Artificial Intelligence Research Society Conference.*
- HLT-NAACL *North American Chapter of the Association for Computational Linguistics - Human Language Technologies Conference.*
- ICAART *International Conference on Agents and Artificial Intelligence.*
- IJCNLP *International Joint Conference on Natural Language Processing.*
- IPM *Information Processing and Management Journal.*
- JADT *Journées Internationales d'Analyse Statistique des Données Textuelles.*
- JCSL *Journal of Computer Speech and Language.*
- JLRE *Journal of Language Resources and Evaluation.*
- JNLE *Journal of Natural Language Engineering.*
- KDD *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*
- LREC *International Conference On Languages Resources and Evaluation.*
- ML *Journal on Machine Learning.*
- MT *Machine Translation Summit.*
- MM *ACM Multimedia Conference.*
- NTCIR *NII Testbeds and Community for Information Access Research Workshop.*
- RANLP *Recent Advances in Natural Language Processing.*
- SEMEVAL *International Workshop on Semantic Evaluation.*
- SIGIR *ACM SIGIR Special Interest Group on Information Retrieval Conference.*
- SMC *International Conference on Systems, Man, and Cybernetics.*
- SPIRE *International Symposium on String Processing and Information Retrieval.*
- TALN *Conférence sur le Traitement Automatique des Langues Naturelles .*
- WIC *International Conference on Intelligent Agent Technology.*
- WWW *World Wide Web Conference.*