

## Ludovic Tanguy

Né le 22 Octobre 1971 à Brest (Finistère)

Nationalité française

CLLE-ERSS

UMR 5263, CNRS & Université Toulouse-Le Mirail

Maison de la Recherche

5, allées Antonio Machado

F-31058 Toulouse cedex 9

France

Mail: [Ludovic.Tanguy@univ-tlse2.fr](mailto:Ludovic.Tanguy@univ-tlse2.fr)

Tél.: +33/0 5-61-50-36-03

Fax: +33/0 5-61-50-46-77

Web: <http://w3.erss.univ-tlse2.fr/membre/tanguy/>

### Table des matières

---

<b>Fonctions exercées</b>	<b>2</b>
<b>Formation</b>	<b>2</b>
<b>Principales responsabilités</b>	<b>2</b>
<b>Enseignement</b>	<b>3</b>
Principales thématiques . . . . .	3
Enseignements actuels (en tant que responsable) . . . . .	3
Enseignements actuels (en tant qu'intervenant) . . . . .	4
Enseignements passés . . . . .	4
Responsabilités universitaires . . . . .	4
Encadrement de mémoires et de stages . . . . .	5
<b>Recherche</b>	<b>7</b>
Activités administratives et responsabilités collectives . . . . .	7
Organisation de conférences . . . . .	7
Comités scientifiques de conférences . . . . .	7
Comités de lecture et expertises . . . . .	7
Jurys de thèse . . . . .	7
Encadrement de thèses . . . . .	7
<b>Publications</b>	<b>9</b>
Liste par thématique . . . . .	9
Liste par type de support . . . . .	9
<b>Projets de recherche financés</b>	<b>13</b>

## Fonctions exercées

---

- 1999– Maître de conférences en sciences du langage:  
– affecté au département Sciences du Langage de l’Université de Toulouse 2 Le Mirail (UTM)  
– membre du laboratoire CLLE (*Cognition, langues, langage, ergonomie*, UMR 5263) anciennement ERSS (*Équipe de Recherche en Syntaxe et Sémantique*, UMR 5610).
- 1998-99 Chercheur en linguistique informatique à l’ISSCO (*Institut pour les études sémantiques et cognitives*), Université de Genève.
- 1996-98 ATER en informatique à la faculté des lettres et sciences humaines Victor Segalen de l’*Université de Bretagne Occidentale*, Brest.
- 1994-96 Chargé de cours en informatique à l’Université de Bretagne Occidentale et à l’Ecole Nationale Supérieure des Télécommunications de Bretagne.
- 1993-96 Allocataire de recherche (allocation du Ministère de l’enseignement supérieur et de la recherche).

## Formation

---

- 2012 Habilitation à Diriger des Recherches, spécialité Linguistique, Université de Toulouse 2. *Complexification des données et des techniques en linguistique : contributions du TAL aux solutions et aux problèmes*  
Soutenue le 11 septembre 2012 devant le jury composé de: A. Condamines (Toulouse 2, promotrice), B. Habert (ENS-Lyon, président), M.-P. Péry-Woodley (Toulouse 2, examinatrice), P. Sébillot (INSA Rennes, examinatrice), M. Valette (INALCO, rapporteur), F. Yvon (Paris Sud, Rapporteur).
- 1997 Thèse de doctorat en informatique de l’université de Rennes 1, mention très honorable<sup>1</sup>. *Traitement automatique de la langue naturelle et interprétation : contribution à l’élaboration d’un modèle informatique de la sémantique interprétative*  
Soutenue le 7 Mai 1997 devant le jury composé de : D. Herman (Rennes 1, président), F. Rastier (CNRS, rapporteur), P. Enjalbert (Caen, rapporteur), J.-P. Barthélemy (ENSTBr, directeur), I. Kanellos (ENSTBr, encadrant), J. Siroux (Rennes 1, examinateur).
- 1993 DEA d’informatique, option Intelligence artificielle et sciences cognitives, mention bien, Université de Rennes 1. *Apprentissage de structures hiérarchiques de représentation de connaissances par algorithmes génétiques*, sous la direction de Serge Garlatti et Dominique Snyers, École Nationale Supérieure de Télécommunications de Bretagne.
- 1993 Diplôme d’ingénieur de l’École Nationale Supérieure de Télécommunications de Bretagne (ENSTBr).
- 1988-90 Classes préparatoires scientifiques, Brest.

## Principales responsabilités

---

- 2008- Responsable de l’axe TAL du laboratoire CLLE
- 2009-11 Membre nommé du Conseil National des Universités, section 7 sciences du langage)
- 2006-08 Directeur du département des sciences du langage, UTM

---

1. L’université de Rennes 1 ne délivre pas de félicitations.

## Enseignement

---

### Principales thématiques

*Traitement automatique des langues* : Principales techniques du TAL, avec un accent sur l'annotation et l'exploitation de corpus. Cet enseignement se décline en cours théoriques, enseignements pratiques et projets encadrés. Public : master ECIL (Ergonomie Cognitive et Ingénierie Linguistique).

*Informatique pour la linguistique* : Outillage informatique à destination des étudiants de sciences du langage. Outils d'annotation et d'interrogation de corpus. Public : licence en sciences du langage.

*Méthodologie en sciences du langage* : Méthodologie de la linguistique empirique (collecte, qualification, annotation et exploitation de données de divers types). Statistique pour la linguistique. Public : masters sciences du langage LiCoCo (Linguistique, Cognition et Communication) et ECIL.

### Enseignements actuels (en tant que responsable)

*Thématiques actuelles de la recherche en TAL (M2 ECIL)* : Ce cours permet aux étudiants de sciences du langage d'appréhender de près les activités de recherche en traitement automatique des langues. Plusieurs chercheurs viennent y présenter leurs travaux actuels (notamment les membres de l'axe TAL que je coordonne au sein du laboratoire CLLE), et nous approfondissons ensuite avec les étudiants les points essentiels des travaux présentés. Volume : 25h.

*Outils informatiques pour la linguistique (L3)* : Ce cours, obligatoire pour la licence de sciences du langage, vise à donner aux étudiants les connaissances nécessaires à une utilisation simple des données langagières au format électronique. J'y présente les principes du codage des textes, ainsi que la théorie et les principaux outils des langages formels (automates, expressions régulières). Une part importante de ce cours est dédiée à la pratique sur machine (manipulation de données, recherche dans des textes). Volume : 25h (également en enseignement à distance).

*Bases de données et statistiques pour la linguistique (M1)* : Ce cours de méthodologie que je partage avec C. Fabre est obligatoire pour tout étudiant de master sciences du langage. Il aborde les principes et les savoir-faire pratiques de la construction et de l'utilisation d'un système de gestion de bases de données pour la collecte, l'organisation et l'exploitation de données langagières. Dans une seconde partie y sont présentés les outils statistiques couramment utilisés pour les approches quantitatives en linguistique (statistique descriptive, mesure de corrélations, tests d'hypothèses). Volume : 25h.

*Informatique pour le traitement automatique des langues 1 et 2 (M1 ECIL)* : Ces deux cours constituent le cœur de l'enseignement de l'informatique adressé aux étudiants du master mention sciences du langage spécialité Ergonomie Cognitive et Ingénierie Linguistique (ECIL). J'y aborde les grandes techniques du traitement automatique des langues (TAL) pour l'exploitation de données (formatage, étiquetage, extraction et calcul). Ces deux cours comprennent une grande partie pratique par le biais de l'apprentissage de la programmation (langage Perl). Volume :  $2 \times 25$ h.

*Projet TAL (M1 ECIL)* : Cet enseignement est l'occasion pour les étudiants du master ECIL de participer activement à un projet de TAL en développant une application en grandeur réelle, depuis sa définition jusqu'à son évaluation. A titre d'exemple, voici quelques-uns des thèmes des projets des années précédentes : étiqueteur morpho-syntaxique complet du français, système de segmentation thématique de textes, outil d'extraction automatique d'opinion, classifieur thématique de documents. Volume : 25h.

*Outils et Méthodes (M2 ECIL)* : L'objectif de ce cours est d'apporter aux étudiants les connaissances nécessaires à une meilleure compréhension des méthodes de développement d'applications informatiques, et plus particulièrement de leurs interfaces graphiques. Il présente les principes de la conception orientée objet et donne lieu à une partie pratique. Volume : 25h.

*Linguistique et TAL (L3)* : Ce cours optionnel propose aux étudiants de L3 une première initiation au traitement automatique des langues sous deux aspects. Le premier est un panorama des

techniques et des analyses utilisées dans les grandes applications du TAL (correction orthographique et grammaticale, recherche d'information, traduction automatique). Le second est une utilisation des techniques du TAL pour l'investigation linguistique, notamment via l'utilisation d'analyseurs et l'exploitation automatique de leurs résultats. Volume: 50h.

*Projet conception (M2 ECIL)*: Dans le cadre du parcours professionnel du master ECIL, les étudiants ont l'occasion de répondre à une véritable demande d'une étude ou d'un développement en traitement automatique des langues. Ils doivent négocier avec un commanditaire et réaliser un travail sur des données réelles. Ce cours nécessite chaque année l'identification en amont du terrain et de la demande. A titre d'exemple, voici quelques commanditaires des années précédentes: Airbus, BEA (Bureau d'Enquêtes et d'Analyses), Ludothèque de Blagnac, Atchik Services. Volume: 25h (encadrement des étudiants).

### **Enseignements actuels (en tant qu'intervenant)**

*Données linguistiques (M1)*: Ce cours (coordonné par C. Fabre) constitue un enseignement obligatoire pour l'ensemble des étudiants du Master recherche. Il fait le point sur la diversité des types de données utilisées en sciences du langage (données issues de l'intuition, des corpus, d'expériences, d'enquêtes), propose une réflexion sur l'adéquation entre l'objet d'étude et le type de données collectées, et encourage le recours à différents modes de collecte complémentaires. Mon intervention porte sur les aspects méthodologiques de caractérisation et d'annotation des données. Plusieurs intervenants viennent présenter leur approche des données dans leur spécialité: dialectologie, sémiotique, phonologie, psycholinguistique. Volume: 8h sur 25.

*Production, gestion et exploitation de documents textuels (M2 ECIL)*: Dans le cadre de ce cours coordonné par C. Fabre et adressé aux étudiants du master ECIL (linguistes et psychologues), mon intervention concerne la problématique de l'utilisation des documents dans le cadre professionnel (cycle de vie), ainsi que les besoins et solutions spécifiques en terme de documents électroniques (production, indexation, structuration, etc.). Volume: 12h sur 50.

### **Enseignements passés**

*Données linguistiques 2 (M2)*: Principes méthodologiques des approches quantitatives en linguistique et nouveaux regards que celles-ci permettent sur des problématiques classiques (12h annuelles, de 2006 à 2011).

*Courants linguistiques (L1)*: Présentation chronologique des principaux courants de pensée de la linguistique, depuis Port-Royal jusqu'à la grammaire générative (50h annuelles, de 1999 à 2001).

*Logique (L2)*: Principes de la logique des propositions et de la logique des prédicats pour la formalisation d'énoncés en langue naturelle (12h annuelles, de 2007 à 2009).

*Enseignements dans d'autres établissements* (en tant qu'ATER et chargé de cours):

- traduction automatique et assistée (Maîtrise LEA et DEA de langues, Université de Bretagne Occidentale – 25h annuelles, de 1997 à 1998);
- analyse de textes assistée par ordinateur (DEA de lettres modernes, Université de Bretagne Occidentale – 15h annuelles, de 1997 à 1998);
- bureautique (DEUG de SHS et Maîtrise de LEA, Université de Bretagne Occidentale, 50h annuelles, de 1997 à 1998);
- logique (élèves-ingénieurs de l'ENSTBr, 25h, 1995);
- langages formels (élèves-ingénieurs de l'ENSTBr, 25h, 1996);
- bases de données relationnelles (DEUG de sciences, Université de Bretagne Occidentale, 50h annuelles – de 1993 à 1995);
- algorithmique et programmation (DEUG de sciences, Université de Bretagne Occidentale – 50h annuelles, de 193 à 1995).

### **Responsabilités universitaires**

- 2012– Membre du CSQ (Collège Scientifique Qualifié), section 07, Université de Toulouse
- 2011– Co-animateur (avec C. Fabre, responsable) de la spécialité ECIL du master sciences du langage

- 2010 Membre du comité de sélection pour un poste de maître de conférences en informatique, université Paul Sabatier, profil : « Langue, représentation des connaissances et apprentissage »
- 2010 Participation à la rédaction de la maquette du master ECIL
- 2009-11 Membre nommé du Conseil National des Universités, section 7
- 2006-10 Co-animateur (avec C. Fabre et M.-P. Péry-Woodley) de la filière TAL (modules spécifiques en L3 et parcours TAL en master)
- 2006-08 Directeur du département des sciences du langage
- 2006- Responsable des stages et des relations avec les entreprises pour le master ECIL
- 2006- Correspondant VAE (Validation des Acquis de l'Expérience) pour le département des sciences du langage
- 2004-06 Directeur adjoint du département des sciences du langage
- 2003- Membre du conseil du département des sciences du langage
- 2001-08 Membre de la commission de spécialistes de l'UTM (section 7)
- 2000-06, 08- Responsable de l'information et de l'orientation des étudiants du département des sciences du langage (UTM). Co-organisation de manifestations (salons du supérieur, journées portes ouvertes, rencontres avec des professionnels)
- 1999-2006 Co-animateur (avec C. Fabre et M.-P. Péry-Woodley) de la licence de sciences du langage mention TAL et de la maîtrise de sciences du langage mention industrie de la langue
- 1999- Responsable informatique du département des sciences du langage (UTM) : gestion du parc, des étudiants vacataires, de l'environnement numérique de travail, etc.
- 1997-98 Vice-président du département d'informatique de la faculté des Lettres et Sciences Sociales de l'Université de Bretagne Occidentale

#### Encadrement de mémoires et de stages

- 2012 – Céline Barès (mémoire de M1 ECIL) : *Repérage de la subjectivité dans des rapports d'incidents et d'accidents aéronautiques*  
 – Laurie Planès (stage de M2 ECIL) : *Etude et comparaison de référentiels aliments*  
 – Camille Bonrepaux (stage de M2 ECIL) : *Internationalisation de BASILES (BAncs SIMulateurs et Logiciels d'Etude de Satellite)*
- 2010 – Sophie Allion (stage de M2 SDL UTM) : *Accès à l'information à partir d'entités nommées : quel bénéfice utilisateur ?*  
 – Simon Leva (mémoire de M1 SDL UTM) : *Extraction d'un lexique bilingue français/tchèque à partir de corpus parallèles alignés*
- 2011 – Camille Bonrepaux (mémoire de M1 SDL UTM) : *La reformulation de requête par les utilisateurs de moteurs de recherche sur le Web : étude de logs*  
 – Simon Leva (mémoire de M2 SDL UTM) : *Caractérisation linguistique des requêtes des utilisateurs d'un moteur de recherche : vers une détection automatique des types de besoins d'information*
- 2010 – Sophie Allion (stage de M2 SDL UTM) : *Accès à l'information à partir d'entités nommées : quel bénéfice utilisateur ?*  
 – Simon Leva (mémoire de M1 SDL UTM) : *Extraction d'un lexique bilingue français/tchèque à partir de corpus parallèles alignés*
- 2009 – Lolla Choucavy (stage de M2 SDL UTM) : *Ontologies pour la recherche d'informations spatiales*  
 – Maud Colléter (mémoire de M1 SDL UTM) : *Vitalis, un programme pour mesurer la vitalité des lexèmes*  
 – Fanny Lalleman (mémoire de M2 SDL UTM) : *Vers une échelle automatique d'importance des références bibliographiques en SHS*  
 – Vanessa Néri (mémoire de M1 SDL UTM) : *Étude des contextes d'apparitions humoristiques des déverbaux, dans le but de les classer automatiquement*

- Nikola Tulechki (mémoire de M2 SDL UTM) : *Parallélisme et Discours : repérage de similarités surfaciques pour l'analyse automatique du discours*
- 2008 – Vanessa Andréani (stage de M2 SDL UTM) : *Étude de la polysémie en anglais pour la désambiguïsation automatique*
- Lionel Fontan (mémoire de M2 SDL UTM) : *Analyse de corpus pour un enseignement de langue de spécialité : aide des outils de TAL*
- David Robin (stage de M2 SDL UTM) : *Préparation et réalisation de l'évaluation du prototype LUISA*
- Nikola Tulechki (mémoire de M1 SDL UTM) : *Étude de la complexité syntaxique sur corpus : des principaux indices à la découverte de dimensions*
- 2007 – Clémentine Adam (mémoire de M1 SDL UTM) : *Traitement automatique de la coréférence pour une application de lexicométrie*
- Julien Bonneau (mémoire de M2 SDL UTM) : *Apports linguistiques pour l'analyse de corpus en sciences humaines : exemples d'applications sur des entretiens semi-directifs*
- Lolla Choucavy (mémoire de M1 SDL UTM) : *Caractérisation des types de textes utilisés en chant choral*
- Yu-Ting Huang (stage de M2 SDL UTM) : *Détection automatique des émotions dans les SMS : étude des indices linguistiques et expressifs*
- Herimanana Zafiharimalala (stage de M2 SDL UTM) : *Modèles mentaux des utilisateurs de la documentation de maintenance*
- 2006 – Coralie Bouffard (mémoire de M1 SDL UTM) : *Exploitation automatique de productions langagières d'aphasiques : le repérage automatique des manifestations de surface du manque du mot*
- Jean-Marc Destabeaux (stage de M2 SDL UTM) : *Extraction terminologique en contexte applicatif*
- Elodie Lafonta (stage de M2 SDL UTM) : *Évaluation et amélioration du système d'archivage de la documentation du domanial*
- 2005 – Marianne Vergez-Couret (mémoire de maîtrise SDL UTM) : *Caractéristiques linguistiques des requêtes en recherche d'information*
- Claudia Ionescu-Ripoll (mémoire de maîtrise SDL UTM) : *Le traitement automatique de la langue roumaine : le cas de la restauration automatique des signes diacritiques*
- 2004 – Herimanana Zafiharimalala (mémoire de maîtrise SDL UTM) : *Classification automatique des noms de jeux et de jouets*
- Aurélia Balesi (mémoire de maîtrise SDL UTM) : *Étude comparative de l'expression de l'opinion selon les genres en vue d'un traitement automatique*
- 2003 – Sophie Nègre (mémoire de maîtrise SDL UTM) : *Les schémas de cooccurrence Verbe/Déverbal : le cas des mots inconnus relevés par Webaffix*
- 2002 – Yalina Alphonse (mémoire de maîtrise SDL UTM) : *Le profilage des textes : apports de l'analyse distributionnelle*
- 2001 – François Provansal (mémoire de maîtrise SDL UTM) : *Étiquetage morphosyntaxique des unités polylexicales*
- 1998 – Marie-Laure Gonthier (stage de maîtrise SDL Paris 7) : *Automatisation de la vérification de traductions*
- 1996 – François Rivière (mémoire de DEA informatique Rennes 1) : *Traitement automatique des langues naturelles et typologie textuelle*
- 1995 – Yannick Prié (mémoire de DEA informatique Rennes 1) : *Contribution à une clarification des rapports entre Informatique et Sémantique Interprétative*

## Recherche

---

### Activités administratives et responsabilités collectives

- 2008– Membre élu du conseil du laboratoire CLLE
- 2008– Responsable de l'axe TAL (traitement automatique des langues) du laboratoire CLLE
- 2005-06 Responsable de l'opération TAL (traitement automatique des langues) du laboratoire ERSS

### Organisation de conférences

- Co-organisateur de la journée d'étude de l'association pour le traitement automatique des langues: « Enseignement du TAL » (2008)
- Membre du comité d'organisation de la conférence TALN 2007
- Organisation d'un tutoriel sur les technologies XML au sein du laboratoire ERSS (2006)
- Journées d'étude de l'opération « Sémantique et corpus » sur *La place des méthodes quantitatives dans le travail du linguiste* (avec Nabil Hathout, 2004)
- Membre du comité d'organisation de la conférence « 39th Annual Meeting of the Association for Computational Linguistics » (ACL'01/EACL'01 Joint Conference, 2001)
- Membre du comité d'organisation et du comité scientifique de la conférence ATLAS'99 (*Arabic Translation and Localization Symposium*, 1999)

### Comités scientifiques de conférences

- « 7<sup>es</sup> Décembrettes » (2010)
- « Journées d'études Toulousaines » (JETOU 2011)
- Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL 2007)
- journée d'étude de l'Association pour le Traitement Automatique des Langues (ATALA): « Articuler les traitements sur corpus » (2005)

### Comités de lecture et expertises

- Session étudiante de la conférence ESSLLI (*European Summer School in Logic, Language and Information*, 2009)
- Expert pour l'Agence Nationale pour la Recherche (ANR): Évaluation d'un projet soumis au programme Blanc (2008)
- Revue des Nouvelles Technologies de l'Information (RNTI, 2009)
- Revue « Les Cahiers de Grammaire » (2003 et 2006)
- Revue « Recherches linguistiques de Vincennes » (2006)
- Colloque JETOU (2005)
- Conférence RECITAL (2005)
- *International Conference on Computational Linguistics* (COLING, 2000 et 2002)

### Jurys de thèse

- 2009 Membre du jury de thèse de Marion Laignelet (sciences du langage): *Analyse discursive pour le repérage automatique de segments obsolescents dans des documents encyclopédiques*, UTM.
- 2008 Membre invité au jury de thèse de Désiré Kompaoré (informatique): *Fusion de systèmes et analyse des caractéristiques linguistiques des requêtes: vers un processus de RI adaptatif*, Université de Toulouse 3.

### Encadrement de thèses

- 2011– Simon Leva en codirection avec Josiane Mothe (IRIT): *Exploitation des citations dans les articles de SHS pour la recherche d'information*
- 2010– Nikola Tulechki: *Détection de signaux faibles dans des bases de données textuelles de rapports d'incidents*

- 2009– Assaf Urieli : *Analyse syntaxique robuste du français par méthode hybride : stratégies d'adaptation aux applications et aux ressources*
- 2007-09 Marion Laignelet en codirection avec Marie-Paule Péry-Woodley (succession de Didier Bourigault) : *Analyse discursive pour le repérage automatique de segments obsolètes dans des documents encyclopédiques*. Soutenue le 25 septembre 2009, mention très honorable avec les félicitations



## Publications

---

### Liste par thématique

Thématiques	Références
Analyse de corpus	[Art1, Art2, Art3, Art10, Act1, Act2, Act3, Act4, Act6, Act7, Act8, Act9, Act10, Com1, Com2, Com3, Com4, Com6, Com12]
Outillage de la linguistique	[Ouv1, Art11, Act18, Rap2, Com13]
Web comme corpus pour la morphologie	[Chap1, Art4, Art5, Art7, Art8, Act11, Act16, Act17, Com7, Com8, Com9, Com10, Com11]
Recherche d'information	[Art6, Act12, Act13, Act14, Act15, Com5]
Sémantique Interprétative	[Art9, Art12, Art13, Act19, Act20, Act21, Act22, Rap3, Com14, Com15, Com16, Com17, Com18, Com19]
Autres travaux	[Act5, Act23, Rap1]

### Liste par type de support

---

## Ouvrage

---

[Ouv1] Ludovic Tanguy and Nabil Hathout. *Perl pour les linguistes : programmes en Perl pour exploiter les données langagières*. TIC et Sciences Cognitives. Hermès sciences publications, 2007.

---

## Chapitre d'ouvrage

---

[Chap1] Nabil Hathout, Fiammetta Namer, Marc Plénat, and Ludovic Tanguy. La collecte et l'utilisation des données en morphologie. In B. Fradin; F. Kerleroux; M. Plénat, editor, *Aperçus de Morphologie*, pages 267–287. Presses Universitaires de Vincennes, 2008.

---

## Articles de revues

---

[Art1] Lydia-Mai Ho-Dac, Cécile Fabre, Marie-Paule Péry-Woodley, Josette Rebeyrolle, and Ludovic Tanguy. On the signalling of multi-level discourse structures. *Discours*, 10, 2012.

[Art2] Ludovic Tanguy, Cécile Fabre, Lydia-Mai Ho-Dac, and Josette Rebeyrolle. Caractérisation des échanges entre patients et médecins : approche outillée d'un corpus de consultations médicales. *Corpus*, 10:137–154, 2011.

[Art3] Marion Laignelet, Marie-Paule Péry-Woodley, and Ludovic Tanguy. Découverte de configurations de traits textuels pour la caractérisation des segments d'obsolescence. *Document Numérique*, 13(3):41–68, 2010.

[Art4] Nabil Hathout, Fabio Montermini, and Ludovic Tanguy. Extensive data for morphology: using the World Wide Web. *Journal of French Language Studies*, 18(1):67–85, 2008.

[Art5] Nabil Hathout and Ludovic Tanguy. WEBAFFIX : une boîte à outils d'acquisition lexicale à partir du Web. *Revue Québécoise de Linguistique*, 32(1):61–84, 2005.

[Art6] Claude Chrisment, Taoufiq Dkaki, Josiane Mothe, Sandra Poulain, and Ludovic Tanguy. Recherche d'information : analyse de différents systèmes réalisant la même tâche. *Ingénierie des systèmes d'information*, 10(1):33–58, 2005.

[Art7] Nabil Hathout, Marc Plénat, and Ludovic Tanguy. Enquête sur les dérivés en -able. *Cahiers de Grammaire*, 28:49–90, 2004.

[Art8] Marc Plénat, Stéphanie Lignon, Nicole Serna, and Ludovic Tanguy. La conjecture de Pichon. *Corpus*, 1:105–150, 2002.

[Art9] Ludovic Tanguy and Théodore Thlivit. Parcours interprétatifs (inter) textuels dans le cadre d'une assistance informatique. *Cahiers de Praxématique*, 33:185–215, 2000.

- [Art10] Josette Rebeyrolle and Ludovic Tanguy. Repérage automatique de structures linguistiques en corpus: le cas des énoncés définitoires. *Cahiers de Grammaire*, 25:153–174, 2000.
- [Art11] Ludovic Tanguy, Susan Armstrong, Pierrette Boullon, and Sabine Lehmann. DiET: Diagnostic et evaluation des systèmes de traitement de la langue naturelle. *Cahiers Langues*, 2:140–150, 1999.
- [Art12] Ioannis Kanellos, Jean Le Dù, François Legras, and Ludovic Tanguy. Assistance informatique à l’interprétation des données en cartographie linguistique: Informatisation anthropocentrée du Nouvel Atlas Linguistique de la Basse-Bretagne. *Géolinguistique*, 8:181–196, 1999.
- [Art13] Michel Schmouchkovitch, Ludovic Tanguy, and Ioannis Kanellos. Code et systèmes de significations dans un cas de délire interprétatif. *Annales Médico-Psychologiques*, 156(6):375–386, 1998.

---

## Actes de conférences

---

- [Act1] Christophe Pimm, Céline Raynal, Nikola Tulechki, Eric Hermann, Grégory Caudy, and Ludovic Tanguy. Natural language processing tools for the analysis of incident and accident reports. In *Proceedings of the International Conference on Human-Computer Interaction in Aerospace (HCI-Aero)*, Brussels, 2012.
- [Act2] Ludovic Tanguy, Basilio Calderone, Franck Sajous, and Nabil Hathout. Authorship attribution: using rich linguistic features when training data is scarce. In *Notebook for PAN at CLEF 2012*, Rome, 2012.
- [Act3] Nikola Tulechki and Ludovic Tanguy. Effacement de dimensions de similarité textuelle pour l’exploration de collections de rapports d’incidents aéronautiques. In *Actes de TALN’12*, Grenoble, 2012.
- [Act4] Stergos Afantenos, Nicholas Asher, Farah Benamara, Myriam Bras, Cécile Fabre, Lydia-Mai Ho-dac, Anne Le Draoulec, Philippe Muller, Marie-Paule Péry-Woodley, Laurent Prévot, Josette Rebeyrolle, Ludovic Tanguy, Marianne Vergez-Couret, and Laure Vieu. An empirical resource for discovering cognitive principles of discourse organisation: the annodis corpus. In *proceedings of LREC*, Istanbul, 2012.
- [Act5] Ludovic Tanguy, Assaf Urieli, Basilio Calderone, Nabil Hathout, and Franck Sajous. A multitude of linguistically-rich features for authorship attribution. In *Notebook for PAN at CLEF 2011*, Amsterdam, 2011.
- [Act6] Lydia-Mai Ho-Dac, Marie-Paule Péry-Woodley, and Ludovic Tanguy. Anatomie des structures énumératives. In *Actes de TALN’2010*, Montréal, 2010.
- [Act7] Pascale Vergely, Anne Condamines, Cécile Fabre, Amélie Josselin-Leray, Josette Rebeyrolle, and Ludovic Tanguy. Analyse linguistique des interactions patient/médecin. In *Actes du colloque Actes éducatifs et de soins*, Nice, 2009.
- [Act8] Ludovic Tanguy and Nikola Tulechki. Sentence Complexity in French: a Corpus-Based Approach. In *Proceedings of of the International Conference on Recent Advances in Intelligent Information Systems (IIS)*, pages 131–145, Krakow, Poland, 06 2009.
- [Act9] Ludovic Tanguy, Fanny Lalleman, Claire François, Philippe Muller, and Patrick Séguéla. RHECITAS: citation analysis of French humanities articles. In *proceedings of Corpus Linguistics*, Liverpool, UK, 2009.
- [Act10] Marie-Paule Péry-Woodley, Nicholas Asher, Patrice Enjalbert, Farah Benamara, Myriam Bras, Cécile Fabre, Stéphane Ferrari, Lydia-Mai Ho Dac, Anne Le Draoulec, Yann Mathet, Philippe Muller, Laurent Prévot, Josette Rebeyrolle, Ludovic Tanguy, Marianne Vergez Couret, Laure Vieu, and Antoine Widlöcher. ANNODIS: une approche outillée de l’annotation de structures discursives. In *Actes de TALN*, page 52, Senlis, 06 2009.
- [Act11] Nabil Hathout, Franck Sajous, and Ludovic Tanguy. Looking for French deverbal nouns in an evolving Web (a short history of WAC). In *proceedings of the Fifth Workshop on Web As Corpus (WAC)*, pages 37–44, San-Sebastian, Spain, 09 2009.

- [Act12] Désiré Kompaore, Josiane Mothe, and Ludovic Tanguy. Combining indexing methods and query sizes in information retrieval in French. In *International Conference on Enterprise Information Systems (ICEIS 2008)*, pages 149–154, Barcelona, Spain, 2008.
- [Act13] Josiane Mothe and Ludovic Tanguy. Linguistic Analysis of Users' Queries: towards an adaptive Information Retrieval System. In *proceedings of the IEEE International Conference on Signal-Image Technology & Internet-Based Systems*, Shanghai, 2007.
- [Act14] Josiane Mothe and Ludovic Tanguy. Unités d'indexation et taille des requêtes pour la recherche d'information en français. In *Actes de la conférence INFORSID 2007 Congrès Informatique des Organisations et Systèmes d'Information et de Décision*, pages 225–241, Perros-Guirec, 2007.
- [Act15] Josiane Mothe and Ludovic Tanguy. Linguistic features to predict query difficulty. In *proceedings of the ACM-SIGIR workshop on Predicting query difficulty - methods and applications*, pages 7–10, Salvador de Bahia, Brazil, 2005.
- [Act16] Ludovic Tanguy and Nabil Hathout. Webaffix : un outil d'acquisition morphologique dérivationnelle à partir du Web. In *Actes de TALN'2002*, page 254, Nancy France, 2002.
- [Act17] Nabil Hathout and Ludovic Tanguy. Webaffix: a tool for finding and validating morphological links on the WWW. In *proceedings of LREC*, Las Palmas, Spain, 2002.
- [Act18] Pierrette Bouillon, Sabine Lehmann, Ian Lewin, David Milward, and Ludovic Tanguy. Discourse data in DiET. In *proceedings of the EACL Workshop on Linguistically Interpreted Corpora (LINC)*, Bergen, Norway, 1999.
- [Act19] Ludovic Tanguy, Susan Armstrong, and Derek Walker. Isotopies sémantiques et vérification de traduction. In *actes de TALN'99*, Cargèse, 1999.
- [Act20] Ludovic Tanguy. Computer-aided language processing: using interpretation to redefine man-machine relations. In *proceedings of the IEEE International Cognitive Technology Conference*, Azty, Japan, 1997.
- [Act21] Ludovic Tanguy and Théodore Thlivit. Pastel: un protocole informatisé d'aide à l'interprétation des textes. In *actes du colloque ILN (Informatique et Langue Naturelle)*, Nantes, 1996.
- [Act22] Ioannis Kanellos, Ludovic Tanguy, and Victor-Hugo Zaldivar-Carillo. Représentation des connaissances et sémantique interprétative. In *actes du Troisième Congrès Européen de Systématique*, Rome, Italie, 1996.
- [Act23] Ioannis Kanellos and Ludovic Tanguy. Du sens interne et externe d'un terme. In *actes des 5<sup>e</sup> journées ERLA-GLAT*, Brest, 1994.

---

## Rapports

---

- [Rap1] Nathalie Aussenac-Gilles and A. Condamines *et al.* Rapport de l'action spécifique STIC "corpus et terminologie" (as 34) rattachée au RTP-DOC (RTP 33). Technical report, RTP-DOC, 2003.
- [Rap2] Ludovic Tanguy, David Milward, and Tibor Kiss. Functional specifications for DiET. Technical report, DiET Deliverable D3.2, LE-4204, 1999.
- [Rap3] Ludovic Tanguy. *Traitement automatique de la langue naturelle et Interprétation: Contribution à l'élaboration d'un modèle informatique de la Sémantique Interprétative*. Thèse de doctorat, Université de Rennes 1, 1997.

---

## Communications sans actes

---

- [Com1] Lydia-Mai Ho-Dac, Cécile Fabre, Marie-Paule Péry-Woodley, Josette Rebeyrolle, and Ludovic Tanguy. High-level discourse structures: topical chains and enumerative structures in a diversified annotated corpus. *Corpus Linguistics Conference*, Birmingham, 2011.
- [Com2] Ludovic Tanguy. Analyse linguistique des contextes de citation dans les publications de shs. *Séminaire ERSSàB*, Bordeaux, 2010.

- [Com3] Cécile Fabre, Anne Condamines, Josette Rebeyrolle, Ludovic Tanguy, and Pascale Vergely. Analyse linguistique des disparités dans les échanges médecin-patient. *Colloque de l'ADELF Epidémiologie sociale et inégalités de santé*, Toulouse, 2009.
- [Com4] Cécile Fabre, Anne Condamines, Josette Rebeyrolle, Ludovic Tanguy, and Pascale Vergely. Analyse linguistique des disparités dans les échanges médecin-patient. *Colloque de l'ADELF Epidémiologie sociale et inégalités de santé*, Toulouse, 2009.
- [Com5] Ludovic Tanguy and Josiane Mothe. Tal et recherche d'information : profilage linguistique des requêtes. *Séminaire IRIT*, Toulouse, 2008.
- [Com6] Josette Rebeyrolle, Didier Bourigault, Cécile Fabre, Amélie Josselin Leray, and Ludovic Tanguy. Un laboratoire d'observation de l'usage du vocabulaire recommandé par les instances officielles françaises. In *Colloque Prescriptions En Langue*, Paris, 11 2007.
- [Com7] Franck Sajous and Ludovic Tanguy. Repérage de créations lexicales sur le Web francophone. In *Journée ATALA sur Le Web comme ressource pour le TAL*, 2006.
- [Com8] Georgette Dal, Stéphanie Lignon, Fiammetta Namer, and Ludovic Tanguy. Toile contre dictionnaires : analyse morphologique en corpus de noms déverbaux concurrents. In *Colloque International sur "Les noms déverbaux"*, Villeneuve d'Ascq, 09 2004.
- [Com9] Ludovic Tanguy and Nabil Hathout. Le français sur internet : recherche automatique de néologismes. *Conférence Multimédia, Internet et Etudes Françaises*, Vancouver, 2003.
- [Com10] Ludovic Tanguy and Nabil Hathout. *Webaffix : un outil pour l'étude la morphologie sur le Web*. *Journées de Morphologie de l'ERSS*, Toulouse, 2002.
- [Com11] Nabil Hathout and Ludovic Tanguy. Autodétection des webnéologismes. *Conférence TALN, Corpus et Web*, Villetaneuse, 2002.
- [Com12] Marie-Paule Péry-Woodley and Ludovic Tanguy. De l'oral à l'écrit : l'exploitation des corpus. *Journées Corpus*, Toulouse, 2002.
- [Com13] David Milward and Ludovic Tanguy. Towards the automatic processing of corpora. *21st Annual Meeting of the German Linguistic Society (DGfS)*, Konstanz, 1999.
- [Com14] Ludovic Tanguy. Assistance informatique à l'interprétation et à l'exploration sémantique des textes. *Journée d'étude de l'ATALA : L'accès sémantique aux banques textuelles*, Paris, 1998.
- [Com15] Michel Schmouchkovitch, Ioannis Kanellos, and Ludovic Tanguy. Délire interprétatif ou liberté de l'interprétation. *XVIIe Journées de l'information psychiatrique*, Strasbourg, 1998.
- [Com16] Ludovic Tanguy. Le logiciel pastel. *Séminaire du groupe Traitement Automatique du Langage du GREYC*, Caen, 1998.
- [Com17] Ludovic Tanguy. Vers une application informatique de la sémantique interprétative : le logiciel pastel et l'assistance à l'interprétation des textes. *Séminaire de Sémantique, Université de Paris IV Sorbonne*, Paris, 1998.
- [Com18] Michel Schmouchkovitch, Sylvie Narquet, and Ludovic Tanguy. Code et systèmes de significations dans un cas de délire interprétatif. *Séminaire de linguistique de l'Université de Bretagne Occidentale*, Brest, 1997.
- [Com19] Ludovic Tanguy and Théodore Thlivit. Informatique et sémantique interprétative : l'ordinateur comme assistant, du texte à l'inter-texte. *Séminaire de linguistique de l'Université de Bretagne Occidentale*, Brest, 1997.

## Projets de recherche financés

---

### **Analyse textuelle des REX (retours d'expérience): apports du traitement automatique des langues à la maîtrise des risques**

Période: 2012-13

Financement: Institut pour la Maitrise Des Risques (IMDR)

Partenaires: CLLE (CNRS / Université de Toulouse 2), CFH (Conseils en Facteurs Humains)

Coordinateur: E. Hermann (CFH)

Rôle: Co-responsable (avec C. Fabre) de l'expertise en TAL

Ce projet vise à faire un état de l'art sur les apports du TAL pour l'exploitation des REX dans les activités à risque (transports, industrie). Il envisage notamment le rôle des différentes techniques (indexation, classification, exploration) des bases documentaires établies pour capitaliser les connaissances dans ces activités professionnelles. Il propose à la fois une comparaison des différentes possibilités offertes par les applications de TAL, et un ensemble de recommandations pour les entreprises concernées.

### **RESOCIT: Citations scientifiques et réseaux sociaux: étude des dynamiques relationnelles impliquées dans la production et la diffusion des publications scientifiques**

Période: 2012-15

Financement: Agence Nationale de la Recherche, programme Blanc (SHS 1)

Partenaires: CLLE (CNRS / Université de Toulouse 2), IRIT (CNRS / Université de Toulouse 3), LERASS (Université de Toulouse 3), LEREPS (Université de Toulouse 1), LISST (CNRS / Université de Toulouse 2)

Coordinateur: B. Milard (LISST)

Rôle: Responsable de la tâche d'analyse des contextes linguistiques des citations

Ce projet propose d'aborder, par le biais d'une méthode mixte (qualitative et quantitative) diverses questions autour de la notion de citation dans les publications scientifiques. Principalement ancré dans la sociologie des sciences, RESOCIT positionne la relation de citation dans le cadre plus vaste des relations entre les chercheurs d'une communauté. Les méthodes déployées vont aborder ces questions par le biais d'entretiens ciblés de chercheurs autour d'une de leurs publications, l'étude du processus de production scientifique et de la diffusion des articles, mais aussi à travers l'étude linguistique des contextes de citations (tâche dans laquelle je suis impliqué).

### **CITRI: Exploitation des citations dans les articles de SHS pour la recherche d'information**

Période: 2011-14

Financement: PRES Université de Toulouse & Région Midi-Pyrénées

Partenaires: CLLE (CNRS / Université de Toulouse 2), IRIT (CNRS / Université de Toulouse 3)

Coordinateur: L. Tanguy (CLLE) & J. Mothe (IRIT)

Rôle: Co-responsable avec J. Mothe, encadrant de la thèse de Simon Leva.

L'objectif est, par le biais d'une analyse linguistique utilisant des techniques de traitement automatique des langues, de mettre en place une méthode opérationnelle permettant d'exploiter les relations de citation entre des publications scientifiques dans le domaine des sciences humaines et sociales afin de faciliter leur exploitation et la recherche d'information par des usagers.

### **Détection de signaux faibles dans des bases de données textuelles de rapports d'incidents**

Période: 2011-14

Financement: Convention CIFRE

Partenaires: CLLE (CNRS / Université de Toulouse 2), Conseil en Facteurs Humains (Toulouse)

Coordinateur: M.-P. Péry-Woodley (CLLE)

Rôle: Co-responsable avec M.-P. Péry-Woodley, encadrant de la thèse de Nikola Tulechki.

L'objectif principal de cette thèse est de proposer des méthodes automatiques et semi-automatiques de détection de signes d'alerte précoces de risques émergents parmi les documents contenus dans de volumineuses bases de données textuelles de rapports d'accidents et incidents.

### **CAAS : Contextual Analysis and Adaptive Search**

Période : 2010-14

Financement : Agence Nationale de la Recherche, programme Contenus et Interactions

Partenaires : CLLE (CNRS / Université de Toulouse 2), IRIT (CNRS / Université de Toulouse 3), LIA (Université d'Avignon)

Coordinateur : J. Mothe (IRIT)

Rôle : Responsable pour CLLE

L'objectif de ce projet est de prendre en compte différents éléments contextuels dans un système de recherche d'information textuelle. Les dimensions du contexte sont : le besoin d'information (requête exprimée et caractéristiques de l'utilisateur), la collection documentaire et les paramètres internes du moteur de recherche.

### **Intermede : Interactions médecin-patient en médecine générale et inégalités sociales de santé – analyses interdisciplinaires**

Période : 2009-11

Financement : Institut de Recherche en Santé Publique

Partenaires : CLLE (CNRS / Université de Toulouse 2), INSERM U558 (Toulouse), LISST (CNRS & UTM), Laboratoire de santé publique et d'épidémiologie (CHU Nantes), LERASS (Université de Toulouse 3) (Toulouse)

Coordinateur : T. Lang (U558)

Rôle : Annotation et exploitation des données

L'objectif du projet INTERMEDE est de comprendre dans quelle mesure les interactions qui se produisent entre le patient et le médecin généraliste dans le cadre de la consultation sont le reflet de certaines inégalités sociales de santé. Il s'appuie sur un corpus de consultations, complété par des entretiens et des questionnaires auprès du patient et du médecin.

### **AnnoDis : Annotation discursive – corpus de référence pour le français et outils d'aide à l'annotation et à l'exploitation**

Période : 2008-10

Financement : Agence Nationale de la Recherche, programme Corpus

Partenaires : CLLE (CNRS / Université de Toulouse 2), IRIT (CNRS / Université de Toulouse 3), GREYC (Université de Caen)

Coordinateur : M.-P. Péry-Woodley (CLLE)

Rôle : Annotation et exploitation des données pour l'approche descendante

Le projet ANNODIS vise la construction d'un corpus de textes annotés au niveau discursif ainsi que le développement d'outils pour l'annotation et l'exploitation de corpus. Les annotations adoptent deux points de vue complémentaires : une perspective ascendante part d'unités de discours minimales pour construire des structures complexes via un jeu de relations de discours ; une perspective descendante aborde le texte dans son entier et se base sur des indices pré-identifiés pour détecter des structures discursives de haut niveau.

### **Rhécitas : Rhétorique des citations dans les articles de SHS**

Période : 2008-09

Financement : TGE Adonis (CNRS)

Partenaires : CLLE (CNRS / Université de Toulouse 2), INIST (CNRS, Nancy), IRIT (CNRS / Université de Toulouse 3), Synapse Développement (Toulouse)

Coordinateur : L. Tanguy (CLLE)

Le projet RHECITAS vise à l'identification automatique des fonctions rhétoriques des citations dans les publications en ligne dans le domaine des SHS. Il fait appel pour ce faire à des techniques de TAL pour identifier et caractériser les différents contextes des appels de citation.

### **ARIEL : Adaptation d'une chaîne de Recherche d'Information sur la base de traitements Linguistiques**

Période : 2004-06

Financement : TCAN (CNRS)

Partenaires: ERSS (CNRS / Université de Toulouse 2), IRIT (CNRS/ Université de Toulouse 3)

Coordinateur: J. Mothe (IRIT) & L. Tanguy (ERSS)

Ce projet vise à étudier différentes techniques et ressources linguistiques pour la définition d'un système de recherche d'information qui s'adapte à l'expression des besoins formulés par l'utilisateur.

#### **WESCONVA: *WEb, Suffixation et CONCurrence des déVerbaux d'Action***

Période: 2003-05

Financement: Institut de Linguistique Française

Partenaires: ERSS (CNRS / Université de Toulouse 2), ATILF (CNRS / Université de Nancy 2), SILEX (CNRS / Université de Lille 3)

Coordinateur: G. Dal (SILEX)

Rôle: Mise en place et gestion de la base de données, organisation de l'annotation manuelle, analyse des données.

Ce projet vise l'étude des phénomènes constructionnels concurrents susceptibles de former des déverbaux d'action (*-age, -ment, -tion*) avec une approche quantitative et contextuelle des données. Il se base sur une comparaison entre le lexique attesté dans des corpus de référence (nomenclatures du TLFi et du Robert Électronique) et sur le Web, pour comprendre les raisons motivant une nouvelle création de déverbal, et déterminer les paramètres qui conditionnent le choix de la forme du déverbal.

#### **YAKWA++: *Interrogation de corpus étiquetés syntaxiquement***

Période: 2001-04

Financement: Institut de Linguistique Française

Partenaires: ERSS (CNRS / Université de Toulouse 2), Bases, Corpus et Langage (CNRS / Université de Nice)

Coordinateur: D. Bourigault, C. Fabre & L. Tanguy (ERSS)

Ce projet vise à développer une interface d'interrogation de corpus analysés syntaxiquement (en l'occurrence par l'analyseur Syntex) pour un usage linguistique.

#### **Étude de l'implantation des termes recommandés**

Période: 2001-02

Financement: Délégation Générale à la Langue Française et aux Langues de France)

Coordinateur: D. Bourigault & L. Tanguy (ERSS)

Partenaire: ERSS (CNRS / Université de Toulouse 2)

Ce projet consiste à mesurer l'impact des recommandations faites par la DGLF concernant les termes techniques à employer dans le domaine de l'économie et des finances (par opposition notamment aux termes anglais ou impropres). L'étude s'est faite exclusivement sur un panel de sites Web représentatifs des différents acteurs concernés (institutions, médias, entreprises, écoles).

#### **IDOL: *IRS-Based Document Localisation***

Période: 1997-99

Financement: Communauté européenne, programme INCO-DC)

Partenaires: ISSCO (Suisse), UMIST (Royaume-Uni), EPOS (France), Universal (Tunisie), IME (Liban)

Coordinateur: R. Belhadj Kacem (EPOS)

Rôle: Responsable du module de vérification de la traduction

Ce projet vise à développer une plate-forme d'aide à la traduction pour les langues anglaise, française et arabe. Il propose un ensemble de modules intégrés (mémoire de traduction, gestionnaire terminologique, vérificateur de traduction).

#### **DiET: *Diagnosis and Evaluation Tools for Natural Language Processing***

Période: 1996-99

Financement: Communauté européenne (LE 4204)

Partenaires: ISSCO (Suisse), IBM (Allemagne), DFKI (Allemagne), SRI (Royaume-Uni), UCD (Irlande), Aérospatiale (France)

Coordinateur : K. Netter (DFKI)

Rôle: Responsable de l'outil de profilage, permettant d'adapter les bancs de test à une application de TAL spécifique

Ce projet vise à développer une méthode outillée pour l'élaboration de bancs de test permettant l'évaluation de différents systèmes de traitement automatique des langues.