

Thomas LAVERGNE  
7, place de la république  
91400 ORSAY  
Tél : 06 32 99 39 30  
Mail : thomas.lavergne@reveurs.org

Né le 7 novembre 1980 (29 ans)  
Nationalité française  
—  
Permis B

## Formation initiale

---

- 2005-2009** ENST & ORANGE LABS.  
**Thèse de Doctorat – Détection des textes non-naturels (soutenue le 03/04/2009).**  
Directeurs : François Yvon, Tanguy Urvoy  
Jury : Isabelle Tellier, Mohand Boughanem, Patrick Gallinari, Jean-Louis Dessales, Romain Vinot
- 1999–2005** IFSIC – UNIVERSITÉ DE RENNES 1.  
**Master recherche d’Informatique (Méthodes formelles) Mention A. Bien (2005) ;**  
**Maîtrise d’Informatique Mention Bien (2004) ;**  
**Licence d’Informatique Mention A. Bien (2003) ;**  
**Deug MASS Section Informatique (2002).**
- 1999** LYCÉE RENÉ-DESCARTES, RENNES.  
**Baccalauréat, Série S, Spécialité science et vie de la terre.**

## Expériences

---

- 2009–2010** LIMSI, UNIVERSITÉ DE PARIS SUD.  
**Postdoc – Modèles de réordonnement pour la traduction statistique.**  
▷ *Graphes de facteurs.* ▷ *Traduction Statistique.* ▷ *Réordonnement.* ▷ *Apprentissage artificiel.* ▷ *Programmation parallèle.*
- 2005–2009** ORANGE LABS, LANNION.  
**Thèse de Doctorat – Détection des textes non-naturels.**  
Directeurs : François Yvon, Tanguy Urvoy  
▷ *Recherche d’information.* ▷ *Apprentissage artificiel.* ▷ *Programmation parallèle.* ▷ *Traitement de gros volumes de données.* ▷ *Traitement automatique des langues.*
- 2005** IRISA, RENNES – ÉQUIPE GALION.  
(5 mois) **Stage de Master – Prédicats algébriques d’entiers.**  
Directeur : Didier Caucal  
Caractérisation des mots infinis présents dans des familles de graphes infinis.

## Publications

---

- 2010** **Efficient learning of sparse conditional random fields for supervised sequence labelling.**  
Nataliya Sokolovska, Lavergne Thomas, Olivier Cappé, François Yvon. In *IEEE Journal of Selected Topics in Signal Processing*
- 2009** **Filtering artificial texts with statistical machine learning techniques.**  
Thomas Lavergne, Tanguy Urvoy, François Yvon. In *Language Resources and Evaluation (LRE)*, special issue on Plagiarism and Authorship Analysis
- Transformation rules and Monte-Carlo sampling : a different approach for statistical paraphrase generation.**  
Jonathan Chevelu, Thomas Lavergne, Yves Lepage, Thierry Moudenc. In *Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics (PACLING)*, pages 230–235, Sapporo, Japan, 2009
- Introduction of a new paraphrase generation tool based on Monte-Carlo sampling.**  
Jonathan Chevelu, Thomas Lavergne, Yves Lepage, Thierry Moudenc. In *Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics (ACL-IJCNLP)*, pages 249–252, Suntec, Singapour, 2009
- Détection des textes non-naturels.**  
Thomas Lavergne. *Phd Thesis, ENST Paris & Orange Labs, 2009*

- 2008**      **Detecting Fake Content with Relative Entropy Scoring.**  
 Thomas Lavergne, Tanguy Urvoy, and François Yvon. In *International Workshop on Plagiarism Analysis, Authorship Identification, and Near-Duplicate Detection (PAN)*, pages 27–31, Patras, Greece, 2008.
- Tracking Web Spam with HTML Style Similarities.**  
 Tanguy Urvoy, Emmanuel Chauveau, Pascal Filoche, and Thomas Lavergne. *ACM Transactions on the Web* 2(1), pages 1–28, 2008.
- Taxonomie de textes peu-naturels.**  
 Thomas Lavergne. In *Journées internationales d'Analyse statistique des Données Textuelles*, volume 2, pages 679–689, Lyon, France, 2008. Presses universitaires de Lyon.
- 2006**      **Tracking Web Spam with Hidden Style Similarity.**  
 Tanguy Urvoy, Thomas Lavergne, and Pascal Filoche. In *Proceedings of the 2nd international workshop on Adversarial information retrieval on the web (AIRWeb)*, pages 25–31, Seattle, USA, 2006.
- Unnatural Language detection.**  
 Thomas Lavergne. In *Rencontres Jeunes Chercheurs en Recherche d'Informations (RJCRI)*, pages 383–388, Lyon, France, 2006.

## Compétences techniques

---

Langages	C, C++, Java, Lua, Python, HTML, Latex
Systèmes	Linux (plusieurs distributions), NetBSD, MacOS X, Windows
Outils	Outils de développement GNU, Eclipse, MS Visual Studio

## Langues et divers

---

Anglais	Lu, écrit, parlé. Anglais technique courant
Espagnol	Niveau scolaire
Loisirs	Photographie, Théâtre, Go, Escrime ancienne, Programmation Organisation des rencontres de théâtre amateur « Trait d'union » (depuis 2000)