

Guillaume Wisniewski

Maitre de conférences Université de Paris

Titres et diplômes

2004–2007 **Doctorat en Informatique**, de l'Université Pierre et Marie Curie.

Titre : *Apprentissage dans les espaces structurés. Application à l'étiquetage de séquences et à la transformation automatique de documents* Directeur de thèse : Patrick Gallinari

Jury :

- Younès Bennani, Professeur, Université Paris 13
- Patrick Gallinari, Professeur, Université Paris 6
- Éric Gaussier, Professeur, Université Grenoble 1
- Marc Sebban, Professeur, Université de Saint-Étienne
- Marc Tommasi, Maitre de conférences, Université Lille 3
- François Yvon, Professeur, Université Paris 11

2003–2004 **Diplôme d'études approfondies (DEA)**, Université Pierre et Marie Curie.

Intelligence Artificielle, Reconnaissance des Formes et Applications (IARFA)

2001–2004 **Diplome d'ingénieur**, Télécom Paris Tech.

Fonctions exercées

depuis sept. 2019 **Maitre de conférences**, Université de Paris et LLF.

2008–2019 **Maitre de conférences**, Université Paris Sud et LIMSI.

2007–2008 **Attaché Temporaire d'Enseignement et de Recherche (ATER)**, Université Pierre et Marie Curie et LIP6.

Jul.–Oct. 2003 **Développeur logiciel**, Webraska Mobile Technologies.

Principales Responsabilités

depuis 2018 **Responsable de l'axe *Linguistique et évaluation des systèmes de traitement automatique des langues*** du GDR LIFT.

depuis 2017 **Secrétaire de l'ATALA** .

- depuis 2017 **Membre du bureau de la CCSU 27 de l'Université Paris Sud.**
- depuis 2014 **Responsable de la 2ème année de la licence Maths-Info.**
- depuis 2011 **Responsable de la 3ème année de la licence MIAGE par apprentissage.**

Expérience d'enseignement

- depuis 2008 **Maitre de conférence, à l'Université Paris Sud.**
 - plus de 2000h d'enseignement, principalement en programmation orientée objet, algorithmique, apprentissage automatique, statistique, théorie de l'information, traitement automatique des langues
 - responsable de la 3ème année de la licence MIAGE par apprentissage (environs 22 étudiants) et la 2ème année de la licence Mathématiques-Informatique (environs 150 étudiants) : coordination pédagogique, recrutement, suivi des étudiants, réalisation des emplois du temps, ...
 - membre des groupes de travail "Licence" et "Master" de l'Université Paris Saclay : définition d'une maquette commune aux différents membres de l'Université Paris Saclay, réflexions sur les passerelles entre établissements, ...
- depuis 2014 **Enseignant, Université Paris Diderot.**
Responsable du cours *Méthodes probabilistes pour le TAL* du M1 de Linguistique Informatique
- depuis 2013 **Enseignant, Télécom Paris Tech,** Intervention en formation continue dans le domaines des *Big Data* (notamment en dans le Mastère Big Data).
- 2011–2014 **Enseignant, École Centrale Paris.**
Responsable du cours d'apprentissage statistique en 3ème année
- 2010–2013 **Enseignant, Institut des Langues Orientales (INALCO).**
Responsable du cours de linguistique de corpus

Participation à des comités de programme

J'ai été relecteur pour les conférences et les journaux suivant :

- 2019 : NAACL (Reviewer + Area Chair), TACL, TALN, ACL, *SEM
- 2018 : NAACL, IJCAI, ICML, COLING, QEAP, ACL, TALN, Odyssey, EMNLP, TACL
- 2017 : EACL, ACL, TALN, EMNLP, NIPS, IWSLT
- 2016 : NAACL, ACL, TALN, InterSpeech, Coling, Rep4NLP

- 2015 : BUCC, AAAI, IATIS, TALN, TAL, TACL, Interspeech, IJCAI, Errare, WPTP
- 2014 : Interspeech, CVSC, TALN, BUCC, TACL, MT Journal, WMT, LNE
- 2013 : TALN, Interspeech, CVSC, TAL
- 2012 : NAACL, TALN
- 2011 : TALN
- 2010 : ICTAI, TALN, CORIA

Expertise

- 2019 **membre de comité de sélection**, poste n° 0588 (Paris 8) et n° 1631 (Paris Sud).
- 2018 **membre de comité de sélection**, poste n° 2228 (IUT d'Orsay).
- 2018 **expert ANR**, appel à projet générique (CE23).
- 2016 **expert ANR**, appel Jeune Chercheur.
- 2015 **expert ANR**, appel Jeune Chercheur.
- 2014 **examineur de la thèse de Ngoc Quang Luong**, intitulée *Word Confidence Estimation and Its Applications in Statistical Machine Translation* (sous la direction de Laurent Besacier et de Benjamin Lecouteux).

Encadrement

Depuis 2008, j'ai encadré ou co-encadré 5 thèses (2 en cours), 3 post-docs et plus de 12 stagiaires (L3, M1 et M2)

- depuis Oct. 2018 **Margot Lacour**, *PhD*, Modélisation de l'acquisition d'une langue étrangère.
- depuis Jan. 2018 **José Carlos Rosales**, *PhD*, Traduction automatique de contenu généré par les utilisateurs.
- 2015-2016 **Ophélie Lacroix**, *PostDoc*, Transfert d'analyseur syntaxique en dépendances entre langues.
- 2013-2014 **Anil Kumar Singh**, *PostDoc*, Estimation de confiance pour la traduction automatique.
- 2012-2013 **Artem Sokolov**, *PostDoc*, Méthodes oracles pour la traduction automatique.
- Jan. 2016-Mar. 2019 **Konstantinos Skianis**, *PhD*, Novel Representations, Regularization & Distances for Text Classification.






- Oct. 2013–Mai 2018 **Elena Knyazeva**, *PhD*, Apprendre par imitation : applications à quelques problèmes d'apprentissage structuré en traitement des langues.
- Oct. 2014–Avr. 2018 **Laurianne Aufrant**, *thèse*, Training parsers for low-resourced languages : improving cross-lingual transfer with monolingual knowledge.

Projets

Depuis 2004, j'ai participé à 12 projets, la plupart avec des partenaires industriels. Les plus récents sont :









- 2018 **Errator**, *plateforme d'aide à l'annotation morpho-syntaxique*, financement : AAP Université Paris Saclay.
- 2018–2021 **FlowCON**, *Apprentissage par renforcement pour le contrôle d'écoulement*, <https://flowcon.cnrs.fr/>, financement : DGA.
- 2016–2020 **ParSiTi**, *Parsing and translating user generated content*, <https://parsiti.github.io/>, financement : ANR.
- 2015–2018 **Odessa**, *Online Diarization Enhanced by recent Speaker identification and Structured prediction Approaches*, <http://www.agence-nationale-recherche.fr/Project-ANR-15-CE39-0010>, financement ANR & SNF.
- 2015–2017 **Papyrus**, *Adaptation au domaine pour l'analyse morpho-syntaxique et l'analyse en dépendances*, financement : programme Rapid de la DGA.
- 2012–2015 **Transread**, *Lecture et interaction bilingues enrichies par les données d'alignement*, <https://transread.limsi.fr/>, financement : ANR.
- 2010–2013 **Trace**, *Traduction Robuste par Analyse et Correction d'Erreurs*, <https://anrtrace.limsi.fr/>, financement : ANR.

Publications





				
Journal international	Conférence internationale	Journal francophone	Conférence francophone	Présentation invitée
8	49	4	20	2







Publications principales : ACL (1 article), EMNLP (3 articles), NAACL (5 articles), EACL (2 articles), COLING (2 articles)

2019







1.  J.C. Rosales, D. Seddah, G. Wisniewski, **Phonetic Normalization for Machine Translation of User Generated Content**, WNUT'19
2.  J.C. Rosales, D. Seddah, G. Wisniewski, **Comparison between NMT and PBSMT Performance for Translating Noisy User-Generated Content**, NoDaLiDa 2019
3.  M.A. Bucci, O. Semeraro, A. Allauzen, G. Wisniewski, L. Cordier and L. Mathelin, **Control of chaotic systems by Deep Reinforcement Learning**, Proceedings of the Royal Society A, 2019
4.  M.A. Bucci, A. Allauzen, L. Cordier, L. Mathelin, O. Semeraro and G. Wisniewski, **Apprentissage d'un modèle dynamique chaotique par un LSTM**, CAP'19
5.  G. Wisniewski, **Combien d'exemples de tests sont-ils nécessaires à une évaluation fiable ? Quelques observations sur l'évaluation de l'analyse morphosyntaxique du français.**, TALN'19
6.  M.A. Bucci, O. Semeraro, A. Allauzen, L. Cordier, G. Wisniewski and L. Mathelin, **Control of a chaotic dynamical system with a deep reinforcement learning approach**, GAMM Annual Meeting.
7.  G. Wisniewski and F. Yvon, **How Bad are PoS Tagger in Cross-Corpora Settings ? Evaluating Annotation Divergence in the UD Project**, NAACL'19
8.  E. Knyazeva, G. Wisniewski and F. Yvon, **"Apprendre à chercher" en Traitement Automatique des Langues : état de l'art et perspectives**, TAL

2018











9.  M.A. Bucci, O. Semeraro, A. Allauzen, L. Cordier, G. Wisniewski and L. Mathelin, **Control-oriented model learning with a recurrent neural network**, Bulletin of the American Physical Society
10.  L. Aufrant and G. Wisniewski, **PanParser : a Modular Implementation for Efficient Transition-Based Dependency Parsing**, PBML'18
11.  L. Aufrant, G. Wisniewski and F. Yvon, **Quantifying training challenges of dependency parsers**, COLING'18
12.  J. Patino, R. Yin, H. Delgado, H. Bredin, A. Komaty, G. Wisniewski, C.

- Barras, N. Evans and S. Marcel, **Low-latency speaker spotting with online diarization and detection**, Odyssey'18
13.  G. Wisniewski and F. Yvon, **Divergences entre annotations dans le projet UD et leur impact sur l'évaluation des performance d'étiquetage morpho-syntaxique**, TALN'18
 14.  J. C. Rosales Núñez and G. Wisniewski **Analyse morpho-syntaxique en présence d'alternance codique**, TALN'18
 15.  M. Apidianaki, G. Wisniewski, A. Cocos and C. Callison-Burch **Automated Paraphrase Lattice Creation for HyTER Machine Translation Evaluation**, NAACL'18
 16.  G. Wisniewski, O. Lacroix and F. Yvon **Automatically Selecting the Best Dependency Annotation Design with Dynamic Oracles**, NAACL'18
 17.  L. Aufrant, G. Wisniewski and F. Yvon **Exploiting Dynamic Oracles to Train Projective Dependency Parsers on Non-projective Trees**, NAACL'18
 18.  G. Wisniewski, **Errorator : a Tool to Help Detect Annotation Errors in the Universal Dependencies Project**, LREC'18



2017


19.  G. Wisniewski, **LIMSI Submission for WMT'17 Shared Task on Bandit Learning**, WMT'17
20.  L. Aufrant and G. Wisniewski, **LIMSI@CoNLL'17 : UD Shared Task**, CoNLL'17
21. G. Adda, J.-M. Crego, O. Lacroix, L. Gong Li, D. Mostefa, V. Popescou, G. Wisniewski, F. Yvon **A multilingual corpus of tweets for domain adaptation**, LIMSI Technical Report
22.  G. Wisniewski, H. Bredin, G. Gelly and C. Barras, **Combining speaker turn embedding and incremental structure prediction for low-latency speaker diarization**, Interspeech'17
23.  É. Bartenlian, M. Lacour, M. Labeau, A. Allauzen, G. Wisniewski and F. Yvon, **Adaptation au domaine pour l'analyse morpho-syntaxique**, TALN'17
24.  G. Wisniewski and O. Lacroix **A Systematic Comparison of Syntactic Representations of Dependency Parsing**, UDW'17
25.  L. Aufrant, G. Wisniewski and F. Yvon **Don't Stop Me Now ! Using Global Dynamic Oracles to Correct Training Biases of Transition-Based Dependency Parsers**, EACL'17

2016







- 26. L. Aufrant, G. Wisniewski **PanParser : a Modular Implementation for Efficient Transition-Based Dependency Parsing**, LIMSI Technical Report
- 27.  L. Aufrant, G. Wisniewski and F. Yvon **Zero-resource Dependency Parsing : Boosting Delexicalized Cross-lingual Transfer with Linguistic Knowledge**, COLING'16
- 28.  A. Allauzen, L. Aufrant, F. Burlot, O. Lacroix, E. Knyazeva, T. Lavergne, G. Wisniewski and F. Yvon **LIMSI@WMT'16 : Machine Translation of News**, WMT'16
- 29.  R. Bawden, G. Wisniewski and H. Maynard, **Investigating gender adaptation for speech translation**, TALN'16
- 30.  L. Aufrant, G. Wisniewski and F. Yvon, **Ne nous arrêtons pas en si bon chemin : améliorations de l'apprentissage global d'analyseurs en dépendances par transition**, TALN'16
- 31.  O. Lacroix, L. Aufrant, G. Wisniewski and F. Yvon, **Apprentissage d'analyseur en dépendances cross-lingue par projection partielle de dépendances**, TALN'16
- 32.  L. Aufrant, G. Wisniewski and F. Yvon, **Cross-lingual alignment transfer : a chicken-and-egg story ?**, MLCL'16
- 33.  O. Lacroix, G. Wisniewski and F. Yvon, **Cross-lingual Dependency Transfer : What Matters ? Assessing the Impact of Pre- and Post-processing**, MLCL'16
- 34.  O. Lacroix, L. Aufrant, G. Wisniewski and F. Yvon, **Frustratingly Easy Cross-Lingual Transfer for Transition-Based Dependency Parsing**, NAACL'16
- 35.  L. Aufrant, G. Wisniewski and F. Yvon **Cross-lingual and supervised models for morphosyntactic annotation : a comparison on Romanian**, LREC'16
- 36.  N. Pécheux, G. Wisniewski and F. Yvon, **Reassessing the Value of Resources for Cross-Lingual Transfer of POS Tagging Models**, LRE'16

2015







- 37.  E. Knyazeva, G. Wisniewski, H. Bredin and F. Yvon, **Structured Prediction for Speaker Identification in TV Series**, Interspeech'15
- 38.  E. Knyazeva, G. Wisniewski and F. Yvon., **Apprentissage par imitation pour l'étiquetage de séquences : vers une formalisation des méthodes d'étiquetage easy-first**, TALN'15




39.  N. Pécheux, A. Allauzen, T. Lavergne, G. Wisniewski and F. Yvon, **Oublier ce qu'on sait, pour mieux apprendre ce qu'on ne sait pas : une étude sur les contraintes de type dans les modèles CRF**, TALN'15

2014







40.  G. Wisniewski, N. Pécheux, S. Gahbiche-Braham, F. Yvon, **Ambiguous Learning through Cross-Lingual Transfer for POS Tagging**, EMNLP'14
41.  G. Wisniewski, N. Pécheux, A. Allauzen and F. Yvon, **LIMSI Submission for WMT'14 QE Task**, WMT'14
42.  G. Wisniewski, N. Pécheux, E. Knyazeva, A. Allauzen, F. Yvon, **Apprentissage partiellement supervisé d'un étiqueteur morpho-syntaxique par transfert cross-lingue**, TALN'14
43.  G. Wisniewski, N. Kübler, F. Yvon, **A Corpus of Machine Translation Errors Extracted from Translation Students Exercises**, LREC'14
44.  A. Sokolov, G. Wisniewski, F. Yvon **Lattice BLEU Oracles in Machine Translation**, ACM Transactions on Speech and Language Processing
45.  G. Wisniewski **Conception et analyse d'un corpus de post-éditions : estimation de qualité et analyse d'erreurs pour la traduction automatique**, 4th Workshop on Translation Quality

2013




46.  G. Wisniewski, A. K. Singh, F. Yvon, **Quality estimation for machine translation : some lessons learned**. Machine Translation
47.  G. Wisniewski, F. Yvon, **Oracle decoding as a new way to analyze phrase-based machine translation**, Machine Translation
48.  N. Kübler, F. Yvon, G. Wisniewski, **Human errors and automatic errors in machine translations. What are the differences?**, ERRARE
49.  A. Singh, G. Wisniewski, F. Yvon, **LIMSI submission for the WMT 13 quality estimation task : an experiment with n-gram posteriors**, WMT
50.  G. Wisniewski, A. Singh, N. Segal, F. Yvon, **Design and analysis of a large corpus of post-edited translations : quality estimation, failure analysis and the variability of post-edition**, MT Summit
51.  G. Wisniewski, **On the predictability of human assessment : when matrix completion meets NLP evaluation**, ACL'13

52.  G. Wisniewski, A. Singh, N. Segal, F. Yvon, **Un corpus d'erreurs de traduction**, TALN'13
53.  G. Wisniewski, F. Yvon, **Fast Large-Margin Learning for Statistical Machine Translation**, CiCLING
54.  G. Wisniewski, F. Yvon, **La tâche de prédiction de la qualité**, Tralogy II



2012


















55.  A. Sokolov, G. Wisniewski, F. Yvon, **Non-linear n -best list reranking with few features**, AMTA
56.  P. Zweigenbaum, G. Wisniewski, M. Dinarelli, C. Grouin, S. Rosset, **Résolution des coréférences dans des comptes rendus cliniques. Une expérimentation issue du défi i2b2/VA 2011**, RFIA
57.  A. Sokolov, G. Wisniewski, F. Yvon, **Computing lattice BLEU Oracle scores for machine translation**, EACL
58.  Y. Zhuang, G. Wisniewski, F. Yvon, **Non-linear models for confidence estimation**, WMT
59.  H. Le, T. Lavergne, A. Allauzen, M. Apidianaki, L. Gong, A. Max, A. Sokolov, G. Wisniewski, F. Yvon, **LIMSI@WMT12**, WMT
60.  M. Apidianaki, G. Wisniewski, A. Sokolov, A. Max, F. Yvon, **WSD for n -best reranking and local language modeling in SMT**, SSST-6

2011





61.  N. Tomeh, M. Turchi, G. Wisniewski, A. Allauzen, F. Yvon, **How Good Are Your Phrases ? Assessing Phrase Quality with Single Class Classification**, IWSLT
62.  A. Allauzen, G. Adda, H. Bonneau-Maynard, J-M. Crego, H. Le, T. Lavergne, A. Max, G. Wisniewski, F. Yvon, A. Sokolov, A. Lardilleux **LIMSI@WMT11**, WMT
63.  C. Grouin, M. Dinarelli, S. Rosset, G. Wisniewski, P. Zweigenbaum, **Co-reference resolution in clinical reports. The LIMSI participation in the i2b2/VA 2011 Challenge**

2010


64.  A. Allauzen, J. M. Crego, I. Durgar El-Kahlout, L. Hai-Son, G. Wisniewski and F. Yvon **LIMSI@IWSLT 2010**, IWSLT
65.  G. Wisniewski, A. Allauzen and F. Yvon **Assessing Phrase-Based Translation Models with Oracle Decoding**, EMNLP

66.  H.-S. Le, A. Allauzen, G. Wisniewski and F. Yvon **Training continuous space language models : some practical issues**, EMNLP
 67.  N. Tomeh, A. Allauzen, F. Yvon and G. Wisniewski **Refining Word Alignment with Discriminative Training**, AMTA
 68.   G. Wisniewski, A. Max, F. Yvon **Recueil et analyse d'un corpus écologique de corrections orthographiques extrait des révisions de Wikipédia**, TALN
 69.  A. Max, G. Wisniewski **Mining Naturally-occurring Corrections and Paraphrases from Wikipedia's Revision History**, LREC
- 2009
70.   A. Allauzen and G. Wisniewski **Modèles discriminants pour l'alignement mot-à-mot**, TAL
- 2008
71.  G. Wisniewski and P. Gallinari, **Experimental Evaluation of the Value of Structure : How to Efficiently Exploit Interdependencies in Sequence Labeling**, ICDM'08
- 2007
72. G. Wisniewski, **Apprentissage dans les espaces structurés Applications à l'étiquetage de séquences et à la transformation automatique de documents**, Thèse de l'Université Pierre et Marie Curie
 73.  G. Wisniewski and P. Gallinari, **Relaxation Labeling for Selecting and Exploiting Efficiently Non-Local Dependencies in Sequence Labeling**, PKDD'07
 74.  G. Wisniewski, F. Maes, L. Denoyer and P. Gallinari, **Probabilistic Model for Structured Document Mapping**, MLDM'07
 75.  G. Wisniewski and P. Gallinari, **From Layout to Semantic : A Reranking Model for Mapping Web Documents to Mediated XML Representations**, RIAO
 76.   G. Wisniewski and P. Gallinari, **Ré-ordonnement pour l'apprentissage de transformations de documents HTML** in EGC'07
 77.   G. Wisniewski, F. Maes, L. Denoyer and G. Patrick, **Modèle probabiliste pour l'extraction de structures dans les documents Web** in Document Numérique
- 2006
78.   G. Wisniewski, L. Denoyer, F. Maes and G. Patrick, **Modèle probabiliste pour l'extraction de structures dans les documents semi-structurés — Application aux documents Web**, CORIA'06

2005

- 79.  G. Wisniewski, L. Denoyer and P. Gallinari, **Classification automatique de documents structurés. Application au corpus d'arbres étiquetés de type XML** in CORIA'05
- 80.  P. Gallinari, G. Wisniewski, F. Maes and D. Ludovic, **Stochastic Models for Document Restructuration** in ECML Workshop on Relational Machine Learning
- 81.  L. Denoyer, G. Wisniewski and P. Gallinari, **Classification automatique de structures arborescentes à l'aide du noyau de Fisher : Application aux documents XML** in Actes du 6ème Congrès Européen de Systémique
- 82.  G. Wisniewski, L. Denoyer and P. Gallinari, **Restructuration automatique de documents dans les corpus semi structurés hétérogènes** in EGC'05

2004

- 83.  L. Denoyer, G. Wisniewski and P. Gallinari, **Document Structure Matching for Heterogeneous Corpora** in Proceedings of SIGIR workshop on Integration of Information Retrieval and Databases