

Marie Candito

Curriculum Vitae

Parcours

Parcours professionnel comme enseignant-chercheur

Depuis 2007	Maître de conférences à l'université Paris Diderot, HC depuis 2018
1998-1999	ATER en sciences du langage, option linguistique informatique, à l'université Lille 3

Parcours professionnel en entreprise

oct. 1999 à août 2007	Ingénieure en linguistique informatique , Sociétés Lexiquest puis Lingway Activité R&D et responsable de projets en classification automatique de documents, moteur de recherche sémantique, interrogation de bases de données en langue naturelle.
-----------------------	---

Formation

sept. 1995 à janv. 1999	Doctorat de Linguistique Théorique, Formelle et Automatique , intitulé « <i>Organisation modulaire et paramétrable de grammaires électroniques lexicalisées. Application au français et à l'italien</i> », sous la direction d'Anne Abeillé, Université Paris 7), Soutenu le 26/01/99, Mention Très Honorable avec félicitations du jury
1994-1995	DEA de Linguistique Théorique, Formelle et Automatique , option Linguistique Informatique, à l'Université Paris 7
1993-1994	Ingénieure à l'Université Roma II, au "Gruppo per l'Elaborazione del Linguaggio Naturale"
1987-1992	Classes préparatoires scientifiques puis Ecole Nationale Supérieure des Techniques Avancées , (ENSTA, Paris). Diplôme d' ingénieur en juin 92.

Responsabilités collectives

- **Directrice adjointe de l'UFR de Linguistique** depuis juillet 2012 (avec Claire Saillard comme directrice). Il s'agit d'une UFR comptant 18 enseignants chercheurs permanents, et à laquelle sont rattachées 3 unités de recherche (LLF, HTL et Alpage (terminé fin 2016)).
- **Membre du conseil de l'UFR de Linguistique**, de sept. 2007 à juin 2012, à nouveau depuis juil. 2016.

Activité pédagogique

Responsable du parcours Linguistique Informatique de la L3 de Sciences du Langage (09/10, 10/11, 11/12 et 17/18). Le cœur de mon activité d'enseignement concerne les différentes formations en Linguistique Informatique de l'université Paris Diderot, en L3 et master de Sciences du langage, parcours « Linguistique Informatique ». **Enseignements principaux à Paris Diderot :**

M2 Linguistique Informatique (15 à 25 étudiants)	▪ Analyse sémantique automatique	Créé avec Pascal Amsili Cours+TD, 28h, 3 ans
	▪ Recherche d'information	Créé, Cours, 12h, 7 ans
M1 Linguistique Informatique	▪ Traduction automatique	Transformé (trad. statistique) Cours + TD, 60h, 9 ans

(15 à 25 étudiants)	▪ Méthodes probabilistes pour le TAL	Préexistant Cours + TD, 60h, 2 ans
	▪ Théorie des langages formels	Préexistant TD, 24h, 2 ans
	▪ Apprentissage automatique pour le TAL	Créé en 14/15 Cours + TD, 60h, 4 ans
L3 Linguistique Informatique (25 étudiants)	▪ Probabilités et statistiques pour le Traitement automatique des langues	Créé en 09/10 Cours + TD, 28h, 4 ans
	▪ Encadrement de projet de Traitement automatique des langues	Préexistant (évolution des projets proposés)
L3 Sciences du langage (60 étudiants)	▪ Grammaires d'unification – LFG	Préexistant Cours + TD, 60h, 2 ans
	▪ Linguistique de corpus	Transformé Cours + TD, 28h, 3 ans
L2 Lettres modernes (120 étudiants)	▪ Linguistique française : syntaxe	Préexistant Cours, 36h, 3 ans

Activité scientifique

1. Encadrement doctoral et scientifique :

- **Enrique Henestroza Anguiano:** *Efficient large-context dependency parsing and correction with distributional lexical resources* (1/9/2009 - soutenue le 27/6/2013, officiellement co-encadrée par Laurence Danlos (Univ. Paris Diderot) et Alexis Nasr (Univ. Aix-Marseille).
- **Marianne Djemaa:** *"Stratégie domaine-par-domaine pour la construction d'un FrameNet du français"* (1/10/2012 - soutenue le 14/06/2017, encadrement 100%, habilitation exceptionnelle à diriger une thèse)
- **Hazem Alsaied** *"Analyse syntaxique et identification d'expressions polylexicales par apprentissage profond"* (1/10/2016, co-encadrement à 50%, avec Mathieu Constant, Université de Lorraine)
- **Vincent Segonne** *"Désambiguïsation automatique de verbes"* (1/10/2017, co-encadrement à 50%, avec Benoît Crabbé, Univ. Paris Diderot)

2. Diffusion et rayonnement

Conférences invitées à des colloques internationaux / Conférences plénières

- Conférence invitée à TLT16 (16th International Workshop on Treebanks and Linguistic Theories), Prague, janv 2018, Annotating and Parsing to semantic frames, <https://ufal.mff.cuni.cz/tlt16/>

Activités éditoriales (expertises, responsabilités de collections...)

- Relectrice pour la revue TAL (en 2014 et 2015), le Journal of Language Modeling (en 2014), le Language Resource and Evaluation Journal (en 2015 et en 2017)
- Relectrice pour diverses conférences nationales et internationales, dont ACL, EMNLP, COLING

Participation à des jurys de thèse (en tant qu'examinatrice):

- **Olivier Michalon**, *Modèles statistiques pour la prédiction de cadres sémantiques*. Univ. d'Aix-Marseille, 4/10/2017
- **Raja Bensalem**, *Construction de ressources linguistiques arabes à l'aide des grammaires de propriétés*. Univ. d'Aix-Marseille et de Sfax (Tunisie), 14/12/2017

- **Sandrine Ollinger**, *Le raisonnement analogique en lexicographie, son informatisation, et son application au Réseau Lexical du Français*. Univ. de Lorraine, 30/03/2014
- **Ophélie Lacroix** *De l'étiquetage syntaxique pour les grammaires catégorielles de dépendance à l'analyse par transition en dépendances non-projective*, Univ. de Nantes, 8/12/2014
- **Assaf Urieli**, *Robust French syntax analysis: reconciling statistical methods and linguistic knowledge in the Talismane toolkit*, Univ. de Toulouse Le Mirail, 17/12/2013

Organisation colloques, conférences, journées d'étude:

- "Workshop chair" à la conférence EMNLP 2011 (conférence avec plus de 400 participants)
- Co-organisatrice de SPMRL 2015 (*Statistical Parsing of Morphologically-Rich languages*). <http://www.spmrl.org/category/spmrl2015.html>

Participation à un réseau de recherche, invitations dans des universités étrangères:

- Séminaire donné le 21/12/2017, Computational Linguistics Seminar, [Heinrich-Heine-Universität Düsseldorf](http://www.heine-universitaet-duesseldorf.de), *From deep syntax to semantic frames*, <https://user.phil.hhu.de/kallmeyer/events/>
- Séminaire donné le 26/4/2017, équipe TALEP, *Représentations sémantiques à la FrameNet : utilisation en apprentissage supervisé*, <http://wikitalep.lif.univ-mrs.fr/doku.php?id=seminaire>
- Présentation à la journée TAL et Intelligence Artificielle, 24/03/2016, *Représentations lexicales pour l'analyse sémantique*, <https://afia.asso.fr/category/nos-activites/les-journees-communes/page/2/>
- séminaire donné le 19/11/2013, équipe [Computational Learning and Computational Linguistics \(CLCL\)](http://www.clcl.univ-geneve.ch) (univ. de Genève) *Strategies for Multiword Expression Analysis and Dependency Parsing*

3. Responsabilités scientifiques

Porteuse du projet ANR ASFALDA (ANR-12-CORD-023): Analyse Sémantique en Frames : Annotation, Lexique, Discours et Automatisation (<https://sites.google.com/site/anrasfalda/>)

- octobre 2012 à juin 2016, budget total : 791 706 euros
- Partenaires : ALPAGE (Univ Paris Diderot / INRIA), Laboratoire de Linguistique Formelle (LLF, Univ Paris Diderot / CNRS), Laboratoire d'Informatique Fondamentale (LIF, Univ Aix-Marseille), CEA-List; MELODI (IRIT, Univ de Toulouse), société Ant'inno.
- Production du **French FrameNet**: <http://asfalda.linguist.univ-paris-diderot.fr/frameIndex.xml>. Il s'agit d'un FrameNet du français ([16], [10]), contenant 105 frames, un lexique de 1112 sens (couples frame/lemme) et environ 16000 instances de frames annotées en corpus (ainsi que les remplisseurs de rôles sémantiques correspondants).

Publications en ouvrages collectifs

- [1] Abeillé A. et **Candito M.**, 2000 : "FTAG : A Lexicalized Tree Adjoining Grammar for French". In Abeillé A. et Rambow O. (éds.) CSLI Publications, Stanford (305-330).

Publications dans revues internationales avec comité de lecture

- [2] Al Saied H., **Candito M.** and Constant M., 2018, A transition-based verbal multiword expression analyzer, in *Phraseology and Multi-word Expressions Series*, Markantonatou S., Ramisch C., Savary A., and Vincze V. eds., Language Science Press.
- [3] Savary A., **Candito M.**, Mititelu V., Bejček E., Cap F., Čéplö S., Cordeiro S. R., Eryiğit G., Giouli V., van Gompel M., HaCohen-Kerner Y., Kovalevskaitė J., Krek S., Liebeskind C., Monti J., Parra C., van der Plas L., QasemiZadeh B., Ramisch C., Sangati F., Stoyanova I. and Vincze V., 2018, The PARSEME multilingual corpus of verbal multiword expressions, in *Phraseology and Multi-word Expressions Series*, Markantonatou S., Ramisch C., Savary A., and Vincze V. eds., Language Sci Press
- [4] Seddah D., **Candito M.** and Henestroza Anguiano E., 2013, A word clustering approach to domain adaptation: Robust parsing of source and target domains, *In Journal of Logic and Computation*, special issue on Grammars, Parsers and Recognizers, 24(2): 395-411.

Publications dans revues nationales avec comité de lecture

- [5] **Candito M.**, 1999 : "Un outil multilingue de construction semi-automatique de Grammaires d'Arbres Adjoints. Application au français et à l'italien." *Traitement Automatique des Langues (TAL)*, 40:2.

Publications dans actes de conférences et ateliers internationaux avec comité de lecture

- [6] **Candito M.**, Guillaume B., Perrier G. and Seddah D. 2017, Enhanced UD Dependencies with Neutralized Diathesis Alternation, Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017), Pisa, Italy, 2017
- [7] Al Saied H., **Candito M.** and Constant M., 2017, The ATILF-LLF System for the PARSEME Shared Task: a transition-based verbal multiword expression tagger. Proceedings of the 13th workshop on multiword expressions (MWE 2017) - shared task, Valencia, Spain, 2017.
- [8] Agata Savary, Carlos Ramisch, Silvio Cordeiro, Federico Sangati, Veronika Vincze, Behrang Qasemi Zadeh, **Marie Candito**, Fabienne Cap, Voula Giouli, Ivelina Stoyanova and Antoine Doucet, 2017, The PARSEME Shared Task on Automatic Identification of Verbal Multi-word Expressions, Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017), Valencia, Spain, 2017.
- [9] Michalon O., Ribeyre C., **Candito M.** and Nasr A. 2016, Deeper syntax for better semantic parsing, Proceedings of the 26th International Conference on Computational Linguistics (**COLING 2016**), Osaka, Japan, 2016.
- [10] Djemaa, M., **Candito, M.**, Muller P. and Vieu L. 2016, Corpus annotation within the French Framenet: methodology and results, Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC), Portorož, Slovenia, 2016.

- [11] Vieu L., Muller P., Djemaa, M. and **Candito, M.**, 2016, A General Framework for the Annotation of Causality Based on FrameNet, Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC), Portorož, Slovenia, 2016.
- [12] Seddah D. and **Candito M.** 2016, Hard Time Parsing Questions: Building a QuestionBank for French, Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC), Portorož, Slovenia, 2016.
- [13] **Candito M.** and Constant M., 2014, Strategies for Multiword Expression Analysis and Dependency Parsing, Proceedings of the 52th Annual Meeting of the Association for Computational Linguistics (**ACL'14**), Baltimore, USA, 2014.
- [14] Ribeyre C., **Candito M.** and Seddah D., 2014, Semi-Automatic Deep Syntactic Annotations of the French Treebank. Proceedings of the 13th International Workshop on Treebanks and Linguistic Theories (TLT13), Tübingen, Germany, 2014.
- [15] **Candito M.**, Perrier G., Guillaume B., Ribeyre C., Fort K., Seddah D. and de la Clergerie E., 2014, Deep Syntax Annotation of the Sequoia French Treebank, Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC), Reykjavik, Iceland, 2014.
- [16] **Candito M.**, Amsili P., Barque L., Benamara, F., Chalendar G., Djemaa M., Haas P., Huyghe R., Mathieu Y., Muller P., Sagot B. & Vieu L., 2014, Developing a French FrameNet: methodology and first results, Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC), Reykjavik, Iceland, 2014.
- [17] Constant M., **Candito M.** and Seddah D., 2013, The LIGM-Alpage architecture for the SPMRL 2013 Shared Task: Multiword Expression Analysis and Dependency Parsing, Proceedings of the Fourth SPMRL Workshop, Seattle, USA, 2013.
- [18] Seddah D., Tsarfaty R., Kübler S., **Candito M.**, Choi J.D., Farkas R., Foster J., Goenaga I., Gojenola Galletebeitia K., Goldberg Y., Green S., Habash N., Kuhlmann M., Maier W., Marton Y., Nivre J., Przepiórkowski A., Roth R., Seeker W., Versley Y., Vincze V., Woliński M., Wróblewska A., 2013, Overview of the SPMRL 2013 Shared Task: A Cross-Framework Evaluation of Parsing Morphologically Rich Languages, Proceedings of the Fourth SPMRL Workshop, Seattle, USA, 2013.
- [19] Seddah D., Sagot B., **Candito M.**, Mouilleron V., Vanessa Combet, 2012, The French Social Media Bank: a Treebank of Noisy User Generated Content, In Proceedings of the 24th International Conference on Computational Linguistics (**COLING 2012**), Mumbai, India
- [20] **Candito M.** et Seddah D., 2012, Effectively long-distance dependencies in French : annotation and parsing evaluation, Proceedings of The 11th International Workshop on Treebanks and Linguistic Theories (TLT'11), Lisbon, Portugal
- [21] Seddah D., Sagot B., **Candito M.**, Virginie Mouilleron et Vanessa Combet, 2012, Building a Treebank of Noisy User Generated Content : The French Social Media Bank, Proceedings of The 11th International Workshop on Treebanks and Linguistic Theories (TLT'11), Lisbon, Portugal
- [22] Henestroza Anguiano E. and **Candito M.**, 2012, Probabilistic lexical generalization for French dependency parsing. In Proceedings of the ACL Workshop SP-SEM-MRL '12.
- [23] Seddah D., Sagot B., **Candito M.**, 2012, The Alpage Architecture at the SANCL 2012 Shared Task: Robust Pre-Processing and Lexical Bridging for User-Generated Content Parsing, in Notes of the first workshop of Syntactic Analysis of Non Canonical Languages (SANCL'2012), colocated with NAACL'2012, Montreal, Canada
- [24] Seddah, D., **Candito M.** et Crabbé B., Henestroza Anguiano E. 2012, Ubiquitous Usage of a French Large Corpus: Processing the Est Republicain Corpus, Actes de la 9ème International Conference on Language Resources and Evaluation (LREC 2012), may 2012, Istanbul, Turkey

- [25] **Candito M.**, Henestroza Anguiano E. et Seddah D., 2011, A Word Clustering Approach to Domain Adaptation : Effective Parsing of Biomedical Texts, Actes de l'International Conference on Parsing Technologies (IWPT 2011), october 2011, Dublin, Ireland
- [26] Henestroza Anguiano E. et **Candito M.**, Resolving Difficult Syntactic Attachments with Parse Correction Empirical Methods on Natural Language Processing (**EMNLP 2011**), july 2011, Edinburg, Scotland
- [27] **Candito M.**, Nivre J., Denis P. et Henestroza Anguiano E., 2010, Benchmarking of Statistical Dependency Parsers for French, Actes de la 23ème "International Conference on Computational Linguistics" (**COLING 2010**), august 2010, Beijing, China
- [28] **Candito M.** et Seddah D., 2010, Parsing word clusters, Actes du NAACL/HLT Workshop on Statistical Parsing of Morphologically Rich Languages (SPMRL 2010), june 2010, Los Angeles, USA
- [29] Seddah D., Chrupała G., Cetinoglu O., van Genabith J. et **Candito M.**, 2010, Lemmatization and Statistical Lexicalized Parsing of Morphologically-Rich Languages. Actes du NAACL/HLT Workshop on Statistical Parsing of Morphologically Rich Languages (SPMRL 2010), june 2010, Los Angeles, USA
- [30] Tsarfaty R., Seddah D., Goldberg Y., Kuebler S., Versley Y., **Candito M.**, Foster J., Rehbein I. and Tounsi L., 2010. Statistical Parsing of Morphologically Rich Languages (SPMRL) What, How and Whither, Actes du NAACL-HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages (SPMRL 2010), Los Angeles, June, 2010
- [31] **Candito M.**, Crabbé B., et Denis P., 2010, Statistical French dependency parsing: treebank conversion and first results, Actes de la 8ème "international conference on Language Resources and Evaluation" (LREC'2010), may 2010, La Valletta, Malta
- [32] Seddah, D., **Candito M.** et Crabbé B., 2009, Crossparser evaluation and tagset variation: a French treebank study. Actes de la 11ème "International Conference on Parsing Technologies" (IWPT'09), octobre 2009, Paris, France
- [33] **Candito M.** et Crabbé B., 2009, Improving generative statistical parsing with semi-supervised word clustering. Actes de la 11ème "International Conference on Parsing Technologies" (IWPT'09) - short paper, octobre 2009, Paris, France
- [34] **Candito M.**, Crabbé B. et Seddah D., 2009, On statistical parsing of French with supervised and semi-supervised strategies. Proceedings of the EACL 2009 workshop "Grammatical Inference for computational linguistics", Athènes, Grèce
- [35] Kahane S., **Candito M.** et De Kercadio Y., 2000 : An alternative description of extractions in TAG, Actes du 5ème *International Workshop on Tree Adjoining Grammars and related frameworks (TAG+5)*, Paris, France
- [36] **Candito M.** et Kahane S., 1998 : Defining DTG derivations to get semantic graphs, Actes du 4ème *International Workshop on Tree Adjoining Grammars and related frameworks (TAG+4)*, Philadelphia, USA
- [37] **Candito M.** et Kahane S., 1998 : Can the derivation tree represent a semantic graph? An answer in the light of Meaning-Text Theory, Actes du 4ème *International Workshop on Tree Adjoining Grammars and related frameworks (TAG+4)*, Philadelphie, USA
- [38] **Candito, M.**, 1998 : Building parallel LTAGs for French and Italian, Actes de la conférence conjointe **ACL/COLING'98**, Montréal, Canada
- [39] **Candito, M.**, 1996 : A principle-based hierarchical representation of LTAGs, Actes de la 16ème "International Conference on Computational Linguistics" **COLING-96**, Copenhagen, Danemark

- [40] **Candito, M.**, 1996 : Generating an LTAG out of a principle-based hierarchical representation, Actes de ACL'students (session étudiante de la 34ème rencontre de l'Association for Computational Linguistics), Santa Cruz, USA

Publications dans actes de conférences nationales avec comité de lecture

- [41] Perrier G., **Candito M.**, Guillaume B., Ribeyre C., Fort K. and Seddah D., 2014, Un schéma d'annotation en dépendances syntaxiques profondes pour le français, in Proceedings of TALN 2014 (Traitement Automatique des Langues Naturelles), pp. 574-579, Marseille, France, 2014.
- [42] **Candito M.** et Seddah D., 2012, Le corpus Sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical, Actes de la conférence Traitement Automatique du Langage Naturel, TALN-2012, Grenoble.
- [43] **Candito M.**, Crabbé B., Denis P. and Guérin F., 2009, Analyse syntaxique du français : des constituants aux dépendances. Actes de la conférence Traitement Automatique du Langage Naturel TALN 2009, Senlis
- [44] Crabbé B. et **Candito M.**, 2008, Expériences d'analyses syntaxique statistique du français. Actes de la conférence Traitement Automatique du Langage Naturel, TALN-2008, Avignon
- [45] **Candito, M.**, Coch, J., 2006 : "Les outils d'extraction d'information développés par le projet WATSON et disponibles dans Lingway KM", atelier technolangue de TALN-2006, Leuven, Belgique
- [46] **Candito, M.**, Kahane, S., 1998 : Une grammaire TAG vue comme une grammaire Sens-Texte précompilée, Actes de la conférence Traitement Automatique du Langage Naturel TALN-98, Paris
- [47] **Candito, M.**, 1996 : Un outil pour la génération automatique de LTAG, Actes de la conférences Traitement Automatique du Langage Naturel TALN-96, Marseille