

Marco Dinarelli

Chargé de recherche *CRCN*

Bâtiment IMAG - 700 avenue Centrale
38400 Saint-Martin-d'Hères
☎ 04 57 42 16 07
✉ marco.dinarelli@univ-grenoble-alpes.fr
🌐 www.marcodinarelli.it
CV à jour au 24/09/2020

Parcours professionnel

- Janvier 2019 à aujourd'hui **Chargé de recherche**, *Laboratoire d'Informatique de Grenoble (LIG)*.
Apprentissage artificiel, Apprentissage profond, Traitement Automatique des Langues, Systèmes Bout-en-bout
- Octobre 2017 - Décembre 2018 **Chercheur invité**, *Laboratoire d'Informatique de Grenoble (LIG)*.
Chercheur invité au LIG, mon futur laboratoire d'accueil après ma mobilité CNRS
- Octobre 2013 - Décembre 2018 **Chargé de recherche**, *LaTTiCe - CNRS, ENS, Paris 3 Sorbonne Nouvelle*.
Modèles probabilistes pour le traitement automatique des langues. Notamment modèles neuronaux pour l'étiquetage de séquences et la détection de chaînes de co-référence
- Mars 2012 - septembre 2013 **Post-doctorat**, *LIMSI-CNRS, Groupe TLP*.
Détection automatique de paraphrase pour l'amélioration de la traduction automatique statistique
- Juin 2010 - février 2012 **Post-doctorat**, *LIMSI-CNRS, Groupe ILES*.
Modèles probabilistes pour la détection automatique d'entités nommées arborées dans des émissions radiophoniques

Formations

- 2006 – 2010 **Thèse en informatique**, *Département d'informatique et de télécommunications de l'Université de Trento (Italie)*.
Traitement automatique de langues et apprentissage artificiel pour l'extraction d'information sémantique de données textuelles non structurées
- 2006 **Diplôme d'ingénieur en informatique**, *Faculté d'ingénierie de l'Université de Pise (Italie)*.
Extraction et fouille d'information, architectures parallèles
- 2003 – 2006 **Master en informatique**, *Faculté d'informatique de l'Université de Pise (Italie)*.
Extraction et fouille d'information
- 1998 – 2003 **Licence en informatique**, *Faculté d'informatique de l'Université de Pise (Italie)*.

Activités de recherche

Mes recherches relèvent de la compréhension du langage naturel, d'un côté, et de la conception de nouveaux modèles neuronaux capables de prendre en compte un contexte d'information long. Ces recherches se déclinent actuellement dans les activités spécifiques suivantes.

Traduction automatique neuronale au niveau du document.

Ces recherches s'inscrivent dans le cadre de la thèse de Lorenzo Lupo au LIG (première année). Elles visent à concevoir des nouveaux modèles neuronaux pour la traduction automatique neuronale qui prennent en compte de façon plus efficace le contexte au niveau du document.

Annotation et analyse du dialecte arabe tunisien *Arabizi*.

Ces recherches sont menées sous forme de co-encadrement de la thèse d'Elisa Gugliotta (deuxième année), en co-tutelle entre l'Université Grenoble Alpes et l'Université de Rome "La Sapienza". Elles visent à développer la première ressource linguistique (un corpus) en *Arabizi* annotée sur plusieurs niveaux, à la fois syntaxiques et sémantiques. Cette ressource sera utilisée à la fois pour des recherches en linguistique et en Traitement Automatique des Langues (TAL).

Compréhension automatique de la parole.

Ces recherches sont menées en partie par moi-même en collaboration avec d'autres chercheurs, en partie sous forme d'encadrement de stages de master. Le but est la conception de nouveaux modèles neuronaux de type *sequence-to-sequence* pour la compréhension automatique de la parole en partant directement du signal audio. Nous essayons également d'intégrer dans ces modèles un contexte global, de façon similaire à ce qui se fait dans la traduction automatique neuronale.

Détection de chaînes de coréférences.

Ces recherches sont menées en partie par moi-même en collaboration avec d'autres chercheurs, en partie sous forme de co-encadrement de thèse. Là encore le but est la conception de modèle neuronaux capable de prendre en compte un contexte important, permettant la désambiguïsation de coréférences très éloignées dans le texte.

Encadrement

Janvier 2020 **Co-encadrement.**

- Présent Thèse financé par la chaire MIAI, portant sur la traduction automatique statistique neuronale au niveau des documents.

Co-encadrement (50%) avec un membre du LIG (Laurent Besacier).

Octobre 2019 **Co-encadrement.**

- Présent Thèse en co-tutelle entre l'Université de Rome "La Sapienza" et l'Université Grenoble Alpes, portant sur la traitement et l'analyse du système d'écriture quasi-oral connu comme *Arabizi*. Co-encadrement (33%) avec un membre du Lidilem (Olivier Kraif) et un membre de l'Université de Rome (Giuliano Mion).

2 publications acceptées dans des colloques, plus une soumission en cours.

Octobre 2016 **Co-Encadrement.**

- Juillet 2020 Thèse financée par le LabEx EFL portant sur la détection des chaînes de coréférences.

Co-encadrement (33%) avec un membre du Lattice (Isabelle Tellier, puis Frédéric Landragin) et un membre de l'INRIA (Eric De La Clergerie).

4 publications acceptées dans des colloques.

Thèse soutenue le 15 juillet 2020.

Octobre 2014 **Co-Encadrement.**

- Octobre 2019 Thèse CIFRE portant sur l'extraction d'information sémantique structurée des données issues des réseaux sociaux.

Co-encadrement (33%) avec un membre du Lattice (Isabelle Tellier, puis Thierry Poibeau) et un responsable de l'entreprise *Synthesio* (Pedro Dias Cardoso).

4 publications acceptées dans des colloques.

Thèse soutenue le 16 octobre 2019.

Octobre 2014 **Co-Encadrement.**

- Novembre 2017 Thèse CIFRE portant sur l'extraction d'information sémantique structurée des données bio-médicales.
Co-encadrement (33%) avec un membre du Lattice (Isabelle Tellier) et un responsable de l'entreprise *Temis* (Christian Lautier).
3 publications acceptées dans des colloques, 1 publication dans une revue nationale.
Thèse soutenue le 23 novembre 2017.

Juin 2015 **Encadrement.**

- août 2019 Étudiants en Stage M2 (5) et M1 (1) portant sur la détection automatique des mentions co-référentes, sur la compréhension automatique de la parole, sur la détection de chaînes de co-références.

Enseignement

Septembre 2019 **Enseignement : TD.**

- 2019 Analyse syntaxique, niveau licence 3, Université Grenoble Alpes. 39 heures.
- décembre 2019
- Février 2015 **Enseignement : TD.**
- mai 2015 Traitement Automatique de Langues, niveau Master 2, Université Pierre et Marie Curie. 40 heures.

Mars 2015 **Enseignement : CM.**

Traitement Automatique de Langues, niveau doctorat, Université Paris 3 Sorbonne Nouvelle. 4 heures.

Productions scientifiques

Toutes les publications sont avec comité de relecture.

Sélection des publications plus importantes ou plus récentes. Pour une liste complète de mes publications : <http://www.marcodinarelli.it>

Revue nationale et internationale

1. Modélisation de dépendances entre étiquettes dans les réseaux neuronaux récurrents
M. Dinarelli, Y. Dupont
Revue TAL (Traitement Automatique des Langues) Volume 58 Numéro 1, France, 2017.
2. Discriminative Reranking for Spoken Language Understanding
Marco Dinarelli, A. Moschitti, G. Riccardi
IEEE Journal of Transactions on Audio, Speech and Language Processing (TASLP), volume 20, issue 2, pages 526 - 539, 2012.
3. Comparing Stochastic Approaches to Spoken Language Understanding in Multiple Languages
S. Hahn, **M. Dinarelli**, C. Raymond, F. Lefèvre, P. Lehnen, R. De Mori, A. Moschitti, H. Ney, G. Riccardi
IEEE Journal of Transactions on Audio, Speech and Language Processing (TASLP), volume 19, issue 6, pages 1569 - 1583, 2010.

Conférences nationale et internationale

1. TARc: Incrementally and Semi-Automatically Collecting a Tunisian Arabish Corpus
E. Gugliotta, **M. Dinarelli**
Language Resources and Evaluation Conference (LREC), Marseille, France, 2020.
2. A Data-Efficient End-to-End Spoken Language Understanding Architecture
M. Dinarelli, N. Kapoor, B. Jabaian, L. Besacier
International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Barcelona, Spain, 2020.

3. Seq2Biseq: Bidirectional Output-wise Recurrent Neural Networks for Sequence Modelling
M. Dinarelli, L. Grobol
 International Conference on Intelligent Text Processing and Computational Linguistics (CICLing),
 La Rochelle, France, 2019.
4. Label-dependency coding in Simple Recurrent Networks for Spoken Language Understanding
M. Dinarelli, V. Vukotic, C. Raymond
 In proceedings of Interspeech, Stockholm, Sweden, 2017.
5. Label-Dependencies Aware Recurrent Neural Networks
 Y. Dupont, **M. Dinarelli**, I. Tellier
 International Conference on Intelligent Text Processing and Computational Linguistics (CICLing),
 Budapest, Hungary, 2017.
6. Improving Recurrent Neural Networks for Sequence Labelling
M. Dinarelli, I. Tellier
 arXiv Technical Report, **2016**.
7. Étude des réseaux de neurones récurrents pour étiquetage de séquences
M. Dinarelli, I. Tellier
 23ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN), Paris, France,
2016.
8. Understanding Social Media Texts with Minimum Human Effort on #Twitter
 T. Tian, I. Tellier, **M. Dinarelli** and P. Dias Cardoso
 Language and the new (instant) media (PLIN), Louvain-la-Neuve, Belgium, **2016**.
9. Domain Adaptation for Named Entity Recognition Using CRFs
 T. Tian, **M. Dinarelli**, I. Tellier and P. Dias Cardoso
 Language Resources and Evaluation Conference (LREC), Portoroz, Slovenia, **2016**.
10. Extraction automatique d'affixes pour la reconnaissance d'entités nommées chimiques
 Y. Dupont, I. Tellier, C. Lautier, **M. Dinarelli**
 Extraction et Gestion des Connaissances, Reims, France, **2016**.
11. Etiquetage morpho-syntaxique de tweets avec des CRF
 T. Tian, **M. Dinarelli**, I. Tellier, P. Cardoso
 Traitement Automatique des Langues Naturelles, Caen, France, **2015**.
12. Data Adaptation for Named Entity Recognition on Tweets with Features-Rich CRF
 T. Tian, **M. Dinarelli**, I. Tellier
 ACL Workshop on Noisy User-generated Text: Twitter Lexical Normalization and Named Entity
 Recognition, Beijing, China, **2015**.
13. Evaluation of Different Strategies for Domain Adaptation in Opinion Mining
 A. Garcia-Fernandez, O. Ferret, **M. Dinarelli**
 Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, **2014**.
14. LIMSI @ WMT'13
 A. Allauzen, N. Pécheux, Q.K. Do, **M. Dinarelli**, T. Lavergne, A. Max, H.S. Le, F. Yvon
 Eighth Workshop on Statistical Machine Translation (WMT), Sophia, Bulgaria, **2013**.

Communication et vulgarisation

Juin 2019 **Atelier**.

Atelier sur la compréhension automatique de la parole (SLU) entre membre du GDR TAL,
 avec le but de collaborer et faire avancer le SLU en France.

- Novembre **Groupe de lecture.**
2018 Animation du groupe de lecture du groupe Getalp du LIG, abordant le sujet de la détection de coréférences.
- Mars 2018 **Séminaire.**
Présentation de mes travaux au LIG pour entamer des collaborations.
- Septembre **Séminaire.**
2016 Conférence à l'École Normale Supérieure de Paris pour présenter mes travaux aux étudiants et à des potentiels collaborateurs.
- Mars 2016 **Séminaire.**
Séminaire au LaTTiCe pour présenter mes travaux aux membres du laboratoire.
- Mai 2015 **Journée des doctorants.**
Participation en tant qu'encadrant et discutant à la journée des doctorants de l'école doctorale de l'Université Paris 3
- Octobre 2014 **Séminaire.**
Séminaire joint du LaTTiCe (4 membres) dans le laboratoire de sciences cognitives et psycholinguistique (LSCP) afin d'établir des collaborations
- Juin 2014 **Journée des doctorants.**
Participation en tant qu'encadrant et discutant à la journée des doctorants de l'école doctorale de l'Université Paris 3
- Mars 2014 **Séminaire.**
Séminaire joint du LaTTiCe (3 membres) dans le laboratoire d'informatique de l'École Normale Supérieure afin d'établir des collaborations