

Lionel NICOLAS

General Informations

- **Date and place of birth** April 17th 1983 in Nice, France
- **Nationality** French
- **Personal status** Single
- **Professional status** Senior researcher in Computational Linguistics, Natural Language Processing
EURAC, European Academy of Bolzano, Italy

Research Interests

The efficiency and linguistic relevance of most *NLP/CL* tools depend directly or indirectly on the quality and coverage of the numeric linguistic resources they rely on. Even for major languages such as Spanish and French, many well known and widely used linguistic resources are still in a precarious state of development. For languages with a smaller speech community, they barely exist.

Such situation is a direct consequence of the ambiguity and irregularities of natural languages making their formalization complex, their manual description labor-intensive and their automatic acquisition difficult.

Nowadays, for linguistic aspects that have consensual frameworks to describe them, one can consider the efforts required to develop linguistic resources as the main limitation. As formalized by Zipf's law, increasing the quality and coverage of a linguistic resource becomes more and more difficult with time since the efforts are always more demanding when compared with the resulting improvements.

In other words, the difficulty for some linguistic levels lies less in how to describe them than in actually achieving a description that has both the coverage and precision required by complex NLP tasks. Owing to these issues, building linguistic resources can take years of constant effort which might too often fail to achieve visible or useful results. Therefore, quick and efficient acquisition and correction of linguistic resources is an open question of considerable interest for the NLP community.

In this context, my research focuses on enhancing the production of symbolic linguistic resources. More specifically, my research objectives focus essentially on:

- practical guidelines for saving efforts when creating or improving linguistic resources,
- the automatized transfer of linguistic knowledge between two closely related languages,
- the automatized acquisition and extension of morphological rules,
- the automatized acquisition and extension of morpho-syntactic wide-coverage lexicons,
- the automatized acquisition and extension of lexicalized grammars,
- the automatized acquisition and extension of phrase translations.

I also have the pleasure to participate in the developments of methods for:

- building learner corpora,
- compiling web corpus,
- linguistic data visualization.

Professional experience

- **Senior researcher in Computational Linguistics, Natural Language Processing**
Since January 2012
EURAC, European Academy of Bolzano, Italy

- **Researcher in Computational Linguistics, Natural Language Processing
+ Assistant teacher in Computer Science**
September 2011 / December 2011
University of Nice Sophia Antipolis, campus Valrose, France

- **Head researcher in Computational linguistics, Natural Language Processing**
February 2011 / June 2011
Molino de Ideas S.A, Madrid, Spain

- **Researcher in Computational Linguistics, Natural Language Processing
+ Assistant teacher in Computer Science**
September 2009 / December 2010
University of Nice Sophia Antipolis, campus Valrose, France

- **Researcher in Computational Linguistics, Natural Language Processing
+ Assistant teacher in Computer Science**
September 2006 / August 2009
University of Nice Sophia Antipolis, campus Valrose, France

- **Personal tutor of English, Spanish, Maths and computer Science**
September 2002 / June 2006
Scolamania school, Nice, France

Academic Background

- **PhD in Computational linguistics / Natural Language Processing,**
"Efficient production of linguistic resources: the Victoria Project"
Under the supervision of Pr. *Jacques FARRÉ*,
RL team, I3S laboratory, University of Nice Sophia-Antipolis,
October 2006 / December 2010.

Grants and contracts

Assistant teacher,
Ministerial research contract,
Mobility grant,

University of Nice Sophia-Antipolis.
French Ministry for Research.
General Council of Alpes Maritimes.

- **Master Degree in Computer Science Research,**
University of Nice Sophia-Antipolis,
October 2004 / July 2006.

Grants

5th year Merit Grant,
Erasmus Grant,

University of Nice Sophia Antipolis.
General Council of Alpes Maritimes.
ETSI2 of Granada (Spain).

2015

- **Correcting OCR errors for German in Fraktur font,** (Nat)
Michel Génèreux, Egon Stemle, Verena Lyding and **Lionel Nicolas**,
Proceedings of **CLIC-it 2014**, Pisa, Italy.
- **Trilingual Learner Corpus illustrating European Reference Levels,** (Nat)
Andrea Abel, Katrin Wisniewski, **Lionel Nicolas**, Adriane Boyd, Jirka Hana and Detmar Meurers,
In **Ricognizioni** – Rivista die Lingue, Letterature e Culture Moderne.

2014

- **InterHist - an interactive visual interface for corpus exploration,** (Int)
Verena Lyding, **Lionel Nicolas** and Egon Stemle,
Proceedings of **LREC 2014**, Reykjavik, Iceland.
- **KoKo: An L1 learner corpus for German,** (Int)
Andrea Abel, Aivars Glaznieks, **Lionel Nicolas** and Egon Stemle,
Proceedings of **LREC 2014**, Reykjavik, Iceland.
- **The MERLIN corpus: Learner language and the CEFR,** (Int)
Adriane Boyd, Jirka Hana, **Lionel Nicolas**, Detmar Meurers, Katrin Wisniewski & al.,
Proceedings of **LREC 2014**, Reykjavik, Iceland.

2013

- **High-accuracy phrase translation acquisition through batte-royale selection,** (Int)
Lionel Nicolas, Egon W. Stemle, Klara Kranebitter and Verena Lyding,
Proceedings of **RANLP 2013**, Hisarya, Bulgaria.
* Acceptance rate 12% (full paper) - Young Researcher Best Paper Award *
- **Open Corpus Interface for Italian Language Learning,** (Int)
Verena Lyding, Claudia Borghetti, Henrik Dittmann, **Lionel Nicolas** and Egon Stemle,
Proceedings of **ICT For Language Learning 2013**, Firenze, Italy.
- **MERLIN: An online trilingual learner corpus empirically grounding the European Reference Levels in authentic learner data,** (Int)
Katrin Wisniewski, Karin Schöne, **Lionel Nicolas**, Chiara Vettori, Adriane Boyd & al.,
Proceedings of **ICT For Language Learning 2013**, Firenze, Italy.
- **A Trilingual Learner Corpus illustrating European Reference Levels,** (Int)
Andrea Abel, **Lionel Nicolas**, Jirka Hana, Barbora Štindlová, Katrin Wisniewski & al.,
Proceedings of **LCR 2013**, Bergen, Norway.
- **A Generic Data Workflow for Building Annotated Text Corpora,** (Nat)
Lionel Nicolas, Egon Stemle, Aivars Glaznieks et Andrea Abel,
Proceedings of **Compiling and using Learner Corpora**, Padova, Italy.
- **StirWaC, Compiling a diverse corpus based on texts from the web for South Tyrolean German,** (Int)
Sarah Schulz, Verena Lyding and **Lionel Nicolas**,
Proceedings of **CL 2013 WAC-8 Workshop**, Lancaster, England.

2012

- **Establishing a Standardised Procedure for Building Learner Corpora,** (Int)
Aivars Glaznieks, Andrea Abel, Verena Lyding, **Lionel Nicolas** and Egon Stemle,
Actes de **LLLC 2012 (Learner Language, Learner Corpora)**, Oulu, Finland.
- **Towards high-accuracy bilingual sequence acquisition from parallel corpora,** (Nat)
Lionel Nicolas, Egon W. Stemle and Klara Kranebitter,
Proceedings of the **Konvens 2012 Workshop Rdalr**, Vienna, Austria.

- **Unsupervised acquisition of concatenative morphology,** *(Int)*
Lionel Nicolas, Jacques Farré and Cécile Darne,
Proceedings of **LREC 2012**, Istanbul, Turkey.
- 2011**

 - **Enriching Lexica through Unsupervised Derivational Rule Acquisition,** *(Int)*
Géraldine Walther and **Lionel Nicolas**,
Proceedings of the **Esslli 2011's International Workshop WoLeR**, Ljubljana, Slovenia.
- 2010**

 - **Efficient production of linguistic resources: the Victoria Project,** *(Int)*
Lionel Nicolas,
PhD Manuscript, RL team, I3S laboratory, University of Nice Sophia-Antipolis, France.
 - **Unsupervised learning of morphology based on frequency-related form occurrence,** *(Int)*
Lionel Nicolas, Jacques Farré and Miguel A. Molinero,
Proceedings of **MorphoChallenge 2010**, Espoo, Finland.
 - **Creating and maintaining language resources: the main guidelines of the Victoria project,** *(Int)*
Lionel Nicolas, Miguel A. Molinero, Benoît Sagot & al.,
Proceedings of the **LRSML workshop of LREC 2010**, Valletta, Malta.
- 2009**

 - **Towards efficient production of linguistic resources: the Victoria Project,** *(Int)*
Lionel Nicolas, Miguel A. Molinero, Benoît Sagot & al.,
Proceedings of **RANLP 2009**, Borovets, Bulgaria.
* Acceptance rate 35% (short paper) *
 - **A morphological and syntactic wide-coverage lexicon for Spanish: the *leffe*,** *(Int)*
Miguel A. Molinero, Benoît Sagot and **Lionel Nicolas**,
Proceedings of **RANLP 2009**, Borovets, Bulgaria.
* Acceptance rate 35% (short paper) *
 - **Producción eficiente de recursos lingüísticos: el proyecto *Victoria*,** *(Nat)*
Lionel Nicolas, Miguel A. Molinero, Benoît Sagot & al.,
Proceedings of **SEPLN 2009**, St. Sebastian, Spain.
 - **Construcción y extensión de un léxico morfológico y sintáctico para el Español: el *leffe*,** *(Nat)*
Miguel A. Molinero, Benoît Sagot and **Lionel Nicolas**,
Proceedings of **SEPLN 2009**, St. Sebastian, Spain.
 - **Trouver et confondre les coupables: un processus sophistiqué de correction de lexique,** *(Nat)*
Lionel Nicolas, Benoît Sagot, Miguel A. Molinero, Jacques Farré and Éric de La Clergerie,
Proceedings of **TALN 2009**, Senlis, France.
* Acceptance rate 27% *
 - **Building a morphological and syntactic lexicon by merging various linguistic resources,** *(Int)*
Miguel A. Molinero, Benoît Sagot and **Lionel Nicolas**,
Proceedings of **NODALIDA 2009**, Odense, Denmark.
 - **FRMG: évolutions d'un analyseur syntaxique TAG du français,** *(Nat)*
Éric De La Clergerie, Benoît Sagot, **Lionel Nicolas** and Marie-Laure Guénot,
Proceedings of **Journée de l'ATALA**, Paris, France.
- 2008**

 - **Mining Parsing Results for Lexical Correction: toward a Complete Correction Process of wide-coverage Lexicons,** *(Int)*
Lionel Nicolas, Benoît Sagot, Miguel A. Molinero, Jacques Farré and Éric de La Clergerie,

Proceedings of **L&TC Lnai best papers 2007**, Poznan, Poland.

*** Best papers Selection ***

- **Computer aided correction and extension of a syntactic wide-coverage lexicon,** *(Int)*
Lionel Nicolas, Benoît Sagot, Miguel A. Molinero, Jacques Farré and Éric de La Clergerie,
Proceedings of **COLING 2008**, Manchester, England.
*** Acceptance rate 24% ***
- **Extensión y corrección semi-automática de léxicos morfo-sintáctico,** *(Nat)*
Lionel Nicolas, Benoît Sagot, Miguel A. Molinero, Jacques Farré and Éric de La Clergerie,
Proceedings of **SEPLN 2008**, Madrid, Spain.

2007

- **Mining Parsing Results for Lexical Corrections,** *(Int)*
Lionel Nicolas, Éric de La Clergerie and Jacques Farré,
Proceedings of **L&TC 2007**, Poznan, Poland.
- **Corrections d'un lexique suggérées par une grammaire,** *(Nat)*
Lionel Nicolas, Éric de La Clergerie and Jacques Farré,
Proceedings of **TALN 2007**, Toulouse, France.

Funding and Grant proposals

Accepted (Acp) - Rejected (Rej)

- **Type :** **Local research project.**
- Date :** **Undetermined**
- Body :** Province of South Tyrol, Legge 14.
- Aims :** MAS: Massen-Methodologie für die Akquirierung von Sprachressourcen
- Contribution:** Main author.
- **Type :** **European research project.** *(Rej)*
- Date :** **January 2016 / January 2019.**
- Body :** European Commission, Horizon 2020.
- Aims :** BLLISS: Bootstrapping Language Learning Interactive Symbiotic Suite
- Contribution:** Main author.
- **Type :** **European Research project.** *(Rej)*
- Date :** **June 2013 / June 2016.**
- Body :** European Commission, Horizon 2020.
- Aims :** "Consensus", Contextual and Semantic Search in Heterogeneous Data Resources.
- Contribution:** Partner.
- **Type :** **National research project.** *(Acp)*
- Date :** **February 2011 / February 2013.**
- Body :** Centro para el Desarrollo Tecnológico Industrial, Spain.
- Aims :** "Plataforma de conocimiento Lingüístico", Spanish linguistic resources.
- Contribution:** Main author.
- **Type :** **Phd grant.** *(Acp)*
- Date :** **Novembre 2010 / Novembre 2013.**
- Body :** University of Nice Sophia Antipolis, France.
- Aims :** Acquisition and correction of lexicalized grammar.
- Contribution:** Main author.
- **Type :** **Post-doc grant.** *(Acp)*
- Date :** **Novembre 2008 / Novembre 2009.**

- Body :** University of Nice Sophia Antipolis, France.
Aims : Acquisition and correction of lexicalized grammar.
Contribution: Main author.
- **Type :** **Local research project,** (Acp)
Date : **Novembre 2008 / Avril 2011.**
Body : Xunta de Galicia, Spain.
Aims : “*Victoria* project”, efficient production of linguistic resources..
Contribution: Main author.

Teaching Experience

Teaching lectures (Mc), theoretical classes (Td) and practical classes (Tp)

| | | | |
|---|--------------------------|---|-------------------|
| 2011-2012 | Assistant Teacher | Campus Valrose, University of Nice Sophia-Antipolis, | 139h Td |
| <u>Licence 1 Math-Info</u> | | | |
| Soutien en Programmation fonctionnelle , introduction to functional programming via Scheme. | | | Td, 13h |
| Introduction à la Programmation fonctionnelle , introduction to functional programming via Scheme. | | | Tp, eq 38h |
| Informatique générale , introduction to fundamental aspects of computer science. | | | Tp, eq 48h |
| <u>Master 1 Miage</u> | | | |
| Compilation et Xml , intermediate classes of XML and related technologies. | | | Td, 40h |
| 2009-2010 | Assistant Teacher | Campus Valrose, University of Nice Sophia-Antipolis, | 96h |
| <u>Licence 1 Math-Info</u> | | | |
| Introduction au web , introduction to Web programming via Html et Css. | | | Tp, eq 3h |
| Introduction à la Programmation Objet , introduction to object-oriented programming via Java. | | | Td, 18h |
| Informatique générale , introduction to fundamental aspects of computer science. | | | Td, 18h |
| <u>Licence 3 Info</u> | | | |
| Compilation , intermediate classes on parsing via Flex/Yacc. | | | Td, 2h |
| Concurrence , intermediate classes on on parallel programming. | | | Td, 16h |
| Stage en entreprise , monitoring of 6-weeks student internship. | | | Tp, eq 6h |
| <u>Master 1 Miage</u> | | | |
| Systèmes distants et Xml , intermediate classes of XML and related technologies. | | | Td, 21h |
| <u>Master 1 Info</u> | | | |
| Programmation objet en C++ , advanced course on object-oriented programming via C++. | | | Td, 12h |
| 2008-2009 | Assistant Teacher | Campus Valrose, University of Nice Sophia-Antipolis, | 57h Td |
| <u>Licence 1 Math-Info</u> | | | |
| Introduction à la Programmation fonctionnelle , introduction to functional programming via Scheme. | | | Tp, eq 24h |
| Introduction à la Programmation C , introduction to programming via C. | | | Tp, eq 28h |
| <u>Master 1 Info</u> | | | |
| Projet de fin d'année , monitoring of students for a 2-months project. | | | eq 5h |
| 2007-2008 | Assistant Teacher | Campus Valrose, University of Nice Sophia-Antipolis, | 66h Td |
| <u>Licence 2 Math-Info</u> | | | |
| Programmation Impérative C , intermediate classes on programming via C. | | | Tp, eq 24h |
| <u>Licence 3 Info</u> | | | |
| Compilation , intermediate classes on parsing via Flex/Yac. | | | Td, 24h |
| Eléments de Génie Logiciel , intermediate classes on software engineering via UML | | | Td, 18h |
| 2006-2007 | Assistant Teacher | Campus Valrose, University of Nice Sophia-Antipolis, | 68h Td |
| <u>Licence 1 Math-Info</u> | | | |

| | |
|--|-------------------|
| Introduction à la Programmation Objet , introduction to object-oriented programming via Java. | Tp, eq 16h |
| <u>Licence 1 Mass</u> | |
| Algorithmique et Calcul Formel , introduction to algorithms and computer algebra. | Tp, eq 12h |
| <u>Licence 2 Math-Info</u> | |
| Programmation Impérative C , intermediate classes on programming via C. | Tp, eq 24h |
| <u>Licence 3 Info</u> | |
| Compilation , intermediate classes on parsing via Flex/Yac. | Tp, eq 16h |

Summer Schools and Formations attended

National (Nat) - International (Int)

| | | |
|-------------|--|--------------|
| 2015 | Workshop on Opinio , UNIBZ, Bolzano, Italy. | <i>(Nat)</i> |
| | Workshop on Intellectual Property Rights , EURAC, Bolzano, Italy. | <i>(Nat)</i> |
| | Workshop on Scientific Research on the Internet , EURAC, Bolzano, Italy. | <i>(Nat)</i> |
| 2014 | Workshop on Convincing presentation techniques , EURAC, Bolzano, Italy. | <i>(Nat)</i> |
| 2013 | Workshop on Project Writing for Experienced Staff , EURAC, Bolzano, Italy. | <i>(Nat)</i> |
| | Workshop on Project Budgeting for Experienced Staff , EURAC, Bolzano, Italy. | <i>(Nat)</i> |
| 2011 | Lisbon Machine Learning School (LXLMS) , Lisbon, Portugal. | <i>(Int)</i> |
| 2009 | European Summer School in Logic, Language and Information (ESSLI) , Bordeaux, France. | <i>(Int)</i> |
| | École d'Été Méthodes Info et Stat en Analyse de Textes (MISAT) , Besançon, France. | <i>(Nat)</i> |
| 2007 | European Summer School in Logic, Language and Information (ESSLI) , Dublin, Ireland. | <i>(Int)</i> |

Conferences and Workshops attended

National (Nat) - International (Int)

| | | |
|-------------|---|--------------|
| 2015 | eLex Conference , Sussex, England | <i>(Int)</i> |
| 2014 | CLARIN ERIC Conference , Wroclaw, Poland. | <i>(Int)</i> |
| | International Conference on Language Resources and Evaluation (LREC) , Reykjavik, Iceland. | <i>(Int)</i> |
| 2013 | CLARIN ERIC Conference , Prague, Czech Republic. | <i>(Int)</i> |
| | Recent Advances in Natural Language Processing (RANLP) , Hisarya, Bulgaria. | <i>(Int)</i> |
| 2012 | CLARIN-Eric Conference , Sophia, Bulgaria. | <i>(Int)</i> |
| | Workshop Recent developments and applications of lexical-semantic resources , Vienna, Austria. | <i>(Nat)</i> |
| | International Conference on Language Resources and Evaluation (LREC) , Istanbul, Turkey. | <i>(Int)</i> |
| 2010 | MorphoChallenge 2010 - Semi-supervised and Unsupervised Analysis , Espoo, Finland. | <i>(Int)</i> |
| | International Conference on Language Resources and Evaluation (LREC) , Valletta, Malta. | <i>(Int)</i> |
| 2009 | Recent Advances in Natural Language Processing (RANLP) , Borovets, Bulgaria. | <i>(Int)</i> |
| | Congreso para el Procesamiento del Lenguaje Natural (SEPLN) , San Sebastian, Spain. | <i>(Nat)</i> |
| | Conférence sur le Traitement Automatique des Langues Naturelles (TALN) , Senlis, France. | <i>(Nat)</i> |
| 2008 | Congreso para el Procesamiento del Lenguaje Natural (SEPLN) , Madrid, Spain. | <i>(Nat)</i> |
| | International Conference on Computational Linguistics (COLING) , Manchester, England. | <i>(Int)</i> |
| | Conférence sur le Traitement Automatique des Langues Naturelles (TALN) , Avignon, France. | <i>(Nat)</i> |
| | Workshop sur le Traitement Automatique des Langues (NATAL) , Nancy, France. | <i>(Nat)</i> |
| 2007 | 3rd Language & Technology Conference (L&TC) , Poznan, Poland. | <i>(Int)</i> |
| | Conférence sur le Traitement Automatique des Langues Naturelles (TALN) , Toulouse, France. | <i>(Nat)</i> |

Langues

- **French** **Mother tongue**
- **English** **Very Fluent**

- **Spanish** **Very Fluent (Dele superior)**
- **Italian** **Very Fluent**
- **German** **Intermediate**
- **Arabic** **Beginner**

Hobbies

Basketball, running, learning languages, traveling and sailing.