



## Office

Department of Linguistics (Office L706)  
2, Rue de Candolle, 1211 Geneva, Switzerland  
Phone: +41 22 379 7332; Fax: +41 22 379 7931  
Violeta.Seretan@unige.ch  
<http://www.latl.unige.ch/personal/vseretan>

# Violeta SERETAN

## PERSONAL DATA

Romanian  
Born January 12, 1976 in Bistrița, Romania  
Married, 1 child  
Languages: English, French, Italian (fluent), Russian (basic), Romanian (native)

## EDUCATION

- PhD, Linguistics • University of Geneva, Switzerland** 2004 – 2008  
Dissertation in Computational Linguistics: “Collocation Extraction Based on Syntactic Parsing”  
Supervisor: Prof. Eric Wehrli  
1<sup>st</sup> mention (summa cum laude, *mention très honorable*)
- MSc, Computer Science • University of Iași, Romania** 1998 – 2000  
Master thesis in Computational Linguistics: “Using Veins Theory's Referential Constraints for Correcting Discourse Structure Analyses” (in Romanian)  
Supervisor: Prof. Dan Cristea  
Final grade: 9.50/10
- BSc, Computer Science • University of Iași, Romania** 1994 – 1998  
Diploma thesis in Computational Linguistics: “An HPSG Approach to Discourse Phenomena” (in Romanian)  
Supervisor: Prof. Dan Cristea  
Final grade: 9.16/10

## RESEARCH INTERESTS

Computational Linguistics (multi-word expressions, lexical resources, corpus-based language processing, terminology, machine translation, syntactic parsing), Formal Languages, Human-Computer Interaction, Computational Geometry

## CURRENT POSITION

**Lecturer (Maître Assistante)** Jul 2008 – present  
LATL – Language Technology Laboratory, Department of Linguistics,  
University of Geneva, Switzerland  
*machine translation, syntactic parsing, lexical resources, collocations*

## RESEARCH EXPERIENCE

### Research/Teaching Assistant, PhD Student

LATL – Language Technology Laboratory, Department of Linguistics,  
University of Geneva, Switzerland

*collocation extraction, concordancing in parallel corpora via  
alignment, corpus-based analysis of collocation morpho-syntactic  
properties, collocation translation, evaluation*

Mar 2002 – Jun 2008

### Summer Intern

FX Palo Alto Laboratory, PARC, Palo Alto, California, U.S.A.

*nominalization, question answering, discourse analysis*

Aug 2005 – Oct 2005

### Research Assistant

Institute of Medicinal Chemistry, University of Lausanne, Switzerland

*computational simulation in chemistry, Web interfaces, 3D molecular  
modelling using VRML, surface rendering*

May 2000 – Jan 2002

### Undergraduate Student Researcher

Faculty of Computer Science, University of Iași, Romania

(Member of the “NLP Group” headed by Prof. Dan Cristea)

*HPSG, discourse theories, anaphora*

1997 – 2000

## PROFESSIONAL EXPERIENCE

### Assistant-manager

“Folk Creation Center Bistrița-Năsăud” (Governmental Organisation),  
Bistrița, Romania

*databases, system administration, desktop publishing*

Sep 1998 – Jan 2000

## TEACHING EXPERIENCE

### Teaching Assistant

Faculty of Arts, University of Geneva, Switzerland

*NLP Laboratory (“Travaux pratiques en relation avec TALN - Traitement  
automatique du langage naturel”)*

Level: 3<sup>rd</sup> year (2002-2005); Master and DEA (2005-2006)

2002 – 2006

### Teaching Assistant

Faculty of Arts, University of Geneva, Switzerland

*UML and Databases (seminar and laboratory for the course “Informatique  
2”)*

Level: 2<sup>nd</sup> year

2004 – 2005

## RESEARCH PROJECTS

Swiss National Science Foundation projects “Analyse multilingue” and  
“Multra”

LATL, Department of Linguistics, University of Geneva, Switzerland

*Contribution to the multilingual development of the Fips symbolic  
parser and of the Its-2 translation system based on syntactic transfer  
with Fips. Development of lexical databases for Fips and Its-2.  
Research on the interrelation between parsing/translation and  
collocations.*

2006 – present

Swiss National Science Foundation project “Fips multilingue”

LATL, Department of Linguistics, University of Geneva, Switzerland

*Contributed to the development of the Fips Romanian parser (lexicon  
development via morphological generation; syntax description;  
corpora consolidation).*

2004 – 2006

- Summer internship project "Denominalizer" 2005  
 FX Palo Alto Laboratory, PARC, Palo Alto, California, U.S.A.  
*Developed a Java program, part of a QA system, for mapping nominal arguments into verbal roles by relying on NomLex (a nominalization resource from New York University) and on selectional restrictions information. Converted NomLex to XML format and implemented Java APIs for NomLex.*
- RUIG-GIAN project "Linguistic Analysis and Collocation Extraction" 2002 – 2004  
 Partners: LATL, University of Geneva; Division of Linguistic Services and Documentation of the WTO (World Trade Organization), Geneva  
*Developed a translation aid tool for the extraction of collocations from corpora, including bilingual concordancing for visualization in parallel corpora, sentence alignment, collocation database management, and Web-based collocate search (in Component Pascal under BlackBox, using ODBC, SQL, and Java).*
- "Web interfaces for molecular modeling" 2000 – 2002  
 Partners: Institute of Medicinal Chemistry and Institute of Computer Science, University of Lausanne, Switzerland  
*Developed part of a server-side program of computational simulation of molecular lipophilicity potential which generates 3D models of molecular structure and surface in VRML (in ANSI C, Perl). Created a client-side web interface with complex interaction functions for the visualization of VRML models (in DHTML).*

## POSTGRADUATE EDUCATION

- Doctoral School in Language Science ("Conférence Universitaires de Suisse Occidentale - Programme doctoral CUSO en Sciences du langage"), Leysin, Switzerland Jun 2006
- Doctoral School in Computer Science ("3eme cycle Romande en Informatique") on "Grid Computing", Anzère, Switzerland Mar 2006
- Doctoral School in Language Science ("Ecole doctorale lémanique en sciences du langage"), University of Geneva, Switzerland May 2004
- Doctoral School in Computer Science ("3eme cycle Romande en Informatique") on "Multimodal and Mobile Interfaces", Anzère, Switzerland Mar 2004
- Doctoral School in Language Science ("Ecole doctorale lémanique en sciences du langage"), University of Geneva, Switzerland May 2003
- SCIE'02: Third Summer Convention on Information Extraction, Frascati, Italy Jul 2002
- Doctoral School in Computer Science ("3eme cycle Romande en Informatique") on "Question/Answering & Web searching", Champéry, Switzerland Mar 2002
- EUROLAN 2001: Summer School on "Creation and Exploitation of Annotated Language Resources", Iași, Romania Aug 2001
- Doctoral School in Computer Science ("3eme cycle Romande en Informatique") on "Autonomous Creatures", Villars, Switzerland Mar 2001
- ESSLLI 2000: European Summer School on Language, Logic and Information, Birmingham, United Kingdom Jul 2000
- EUROLAN 1999: European Summer School on Lexical Semantics and Multilinguality, Iași, Romania Jul 1999
- EUROLAN 1997: European Summer School on Corpus Linguistics, Tușnad, Romania Jul 1997

## CONFERENCE ORGANISATION (CO-CHAIR)

Student Research Workshop at 45<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (ACL-SRW'07), Prague, Czech Republic, June 26, 2007

Student Research Workshop at 11<sup>th</sup> Conference of the European Chapter of the Association for Computational Linguistics (EACL 2006), Trento, Italy, April 6, 2006

## SCIENTIFIC REVIEWING

ACL/IJCNLP 2009 Workshop "Multiword Expressions: Identification, Interpretation, Disambiguation and Applications (MWE 2009)", Singapore), August 6, 2009

Natural Language Engineering journal, 2008

Language Resources and Evaluation journal, 2008

ConsILR 2008 Workshop "Romanian Linguistic Resources and Tools for Natural Language Processing", Iași, Romania, November 19-20, 2008

LREC 2008 Workshop "Towards a Shared Task for Multiword Expressions (MWE 2008)", Marrakech, Morocco, June 1, 2008

ConsILR 2007 Workshop "Romanian Linguistic Resources and Tools for Natural Language Processing", Iași, Romania, December 14-15, 2007

Recent Advances in Natural Language Processing (RANLP-07), Borovetz, Bulgaria, September 27-29, 2007

Recent Advances in Natural Language Processing (RANLP-07) Workshop on "Acquisition and Management of Multilingual Lexicons", Borovetz, Bulgaria, September 30, 2007

Doctoral Consortium at the 8th EUROLAN Summer School, Iași, Romania, July 30 - August 2, 2007

ACL 2007 Workshop "A Broader Perspective on Multiword Expressions", Prague, Czech Republic, June 28, 2007

Computational Linguistics journal, Special Issue on Prepositions, December 2007, Vol. 33, No. 4

COLING/ACL Workshop on Multiword Expressions, Sydney, Australia, July 23, 2006

EACL 2006 Workshop "4th International workshop on ROBust Methods in Analysis of Natural language Data (ROMAND 2006)", Trento, Italy, April 6, 2006

## TALKS

Invited talk: "Collocation extraction with a multilingual parser: Demonstration for Greek", Department of Philology, Linguistics Division, University of Patras, Greece, March 31, 2009

"The LATL collocation extraction tool", *Conférence Universitaires de Suisse Occidentale - Programme doctoral CUSO en Sciences du langage*, Leysin, Suisse, June 15, 2006

"Extraction de collocations", system demonstration, *Le temps d'une découverte*, Service de Presse, Information et Publications, University of Geneva, Switzerland, November 15, 2006

"Decoding the predicate-argument structure of nominalizations", *Summer Intern Presentation*, FX Palo Alto Laboratory, Palo Alto, California, U.S.A., October 26, 2005

"Some questions on hybrid collocation extraction", *Rencontre Erasmus CUI-LATL*, University of Geneva, Switzerland, July 7, 2005

"Beyond word pairs: extracting long multi-word units from parsed text using a method of pair composition", *BAAL Corpus Linguistics SIG/OTA Workshop on Identifying and Researching Multi-Word Units*, Oxford University, U.K., April 21, 2005

"Syntactic-based collocation extraction from parallel corpora and from the Web", *Ecole doctorale lémanique en sciences du langage*, University of Geneva, Switzerland, May 6, 2004

"Collocation extraction in aligned corpora", *Ecole doctorale lémanique en sciences du langage*, University of Geneva, Switzerland, May 2, 2003

"Analyse linguistique et extraction des collocations", *Séminaire de recherche de LATL*, University of Geneva, Switzerland, December 5, 2002

Invited talk: "Discourse structure and references: using referential constraints to correct arbitrary discourse analyses", *Open Day on Discourse Representation and Analysis*, EPFL, Lausanne, Switzerland, May 22, 2001

## PUBLICATIONS

### Dissertation

Violeta Seretan. 2008. *Collocation Extraction Based on Syntactic Parsing*. Ph.D. thesis, University of Geneva.

### Book chapters

Violeta Seretan and Eric Wehrli (forthcoming). "Context-sensitive look-up in electronic dictionaries". In Rufus H. Gouws, Ulrich Heid, Wolfgang Schweickard, Herbert Ernst Wiegand, editors, *Dictionaries. An international encyclopedia of lexicography. Supplementary volume: Recent developments with special focus on computational lexicography, Handbooks of Linguistics and Communications Science*. Walter de Gruyter, Berlin/New York.

Violeta Seretan, Luka Nerima, and Eric Wehrli. 2004. "Multi-word collocation extraction by syntactic composition of collocation bigrams". *Recent Advances in Natural Language Processing III: Selected Papers from RANLP 2003*, Nicolas Nicolov et al. eds., 91–100. Amsterdam & Philadelphia: John Benjamins.

### Journal publications

Violeta Seretan (forthcoming). Extraction de collocations et leurs équivalents de traduction à partir de corpus parallèles ('Extracting collocations and translation equivalents from parallel corpora'). *TAL*, 50(1). In French.

Violeta Seretan and Eric Wehrli. 2008. Multilingual collocation extraction with a syntactic parser. *Language Resources and Evaluation*, 43(1), 71–85. DOI: 10.1007/s10579-008-9075-7.

Luka Nerima, Violeta Seretan, and Eric Wehrli. 2006. Le problème des collocations en TAL. *Nouveaux cahiers de linguistique française*, 27(2006):95–115.

### Peer-reviewed conference publications

Violeta Seretan (forthcoming). An integrated environment for extracting and translating collocations. In *Proceedings of The Fifth Corpus Linguistics Conference*, Liverpool, U.K.

Eric Wehrli, Violeta Seretan, Luka Nerima, and Lorenza Russo. 2009. Collocations in a rule-based MT system: A case study evaluation of their translation adequacy. In *Proceedings of The 13th Annual Meeting of the European Association for Machine Translation*, pages 128–135, Barcelona, Spain.

Athina Michou and Violeta Seretan. 2009. A tool for multi-word expression extraction in Modern Greek using syntactic parsing. In *Proceedings of the Demonstrations Session at EACL 2009*, pages 45–48, Athens, Greece.

Violeta Seretan and Eric Wehrli. 2007. Collocation translation based on sentence alignment and parsing. In *Actes de la 14e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2007)*, pages 401–410, Toulouse, France. **(Best Paper Award.)**

Vincenzo Pallotta, Violeta Seretan, and Marita Ailomaa. 2007. User requirements analysis for Meeting Information Retrieval based on query elicitation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL 2007)*, pages 1008–1015, Prague, Czech Republic.

Violeta Seretan and Eric Wehrli. 2006. Accurate collocation extraction using a multilingual parser. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING/ACL 2006)*, pages 953–960, Sydney, Australia.

Violeta Seretan and Eric Wehrli. 2006. Multilingual collocation extraction: Issues and solutions. In *Proceedings of COLING/ACL Workshop on Multilingual Language Resources and Interoperability*, pages 40–49, Sydney, Australia.

Violeta Seretan. 2005. Induction of syntactic collocation patterns from generic syntactic relations. In *Proceedings of Nineteenth International Joint Conference on Artificial Intelligence (IJCAI 2005)*, pages 1698–1699, Edinburgh, Scotland.

Violeta Seretan, Luka Nerima, and Eric Wehrli. 2004. A tool for multi-word collocation extraction and visualization in multilingual corpora. In *Proceedings of the Eleventh EURALEX International Congress (EURALEX 2004)*, pages 755–766, Lorient, France.

Violeta Seretan, Luka Nerima, and Eric Wehrli. 2004. Using the Web as a corpus for the syntactic-based collocation identification. In *Proceedings of International Conference on Language Resources and Evaluation (LREC 2004)*, pages 1871–1874, Lisbon, Portugal.

Violeta Seretan, Luka Nerima, and Eric Wehrli. 2003. Extraction of multi-word collocations using syntactic bigram composition. In *Proceedings of the Fourth International Conference on Recent Advances in NLP (RANLP-2003)*, pages 424–431, Borovets, Bulgaria.

Luka Nerima, Violeta Seretan, and Eric Wehrli. 2003. Creating a multilingual collocation dictionary from large text corpora. In *Proceedings of the Research Notes Session of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL'03)*, pages 131–134, Budapest, Hungary.

Violeta Seretan. 2002. Discourse analysis correction using anaphoric cues. In *Proceedings of the 2nd Workshop on RObust Methods in Analysis of Natural language Data (ROMAND 2002)*, pages 79–86, Frascati, Italy.

Violeta Seretan and Dan Cristea. 2002. The use of referential constrains in structuring discourse. In *Proceedings of The Third International Conference on Language Resources and Evaluation (LREC 2002)*, pages 1231–1238, Las Palmas, Spain.

## **Edited proceedings**

Chris Biemann, Violeta Seretan, and Ellen Riloff, editors. 2007. *Proceedings of the ACL 2007 Student Research Workshop*. Association for Computational Linguistics, Prague, Czech Republic.

Sebastian Padó, Jonathon L Read, and Violeta Seretan, editors. 2006. *Proceedings of the Student Research Workshop at EACL 2006*. Association for Computational Linguistics, Athens, Greece.