

Koning Hollewijnstraat 27
1336HN Almere

Phone: +31 (0)36 521 4885

E-mail: m.b.villada.moiron@rug.nl

WWW: odur.let.rug.nl/~begona

Date of birth

July 11, 1972

Citizenship

Spanish

RESEARCH
INTERESTS

Lexical acquisition, statistical natural language processing, computational lexicography, grammar development, corpus linguistics

EDUCATION

2000 - 2005 Ph.D., University of Groningen, Computational Linguistics
Title: "Data-driven identification of fixed expressions and their modifiability"
1998 - 2000 M.Sc., Trinity College Dublin, Computational Linguistics
Thesis Title: "Status of Galician and Spanish Pronominal Clitics."
1993 - 1997 B.A., University of Santiago de Compostela, English Philology
1990 - 1993 Diploma, University of Santiago de Compostela, Primary School Teaching

ACADEMIC
EXPERIENCE

Project Member

05/2005 - 08/2007 Identification and lexical Representation of Multiword Expressions (IRME)
Research on hybrid probabilistic models to identify multiword expressions (MWES) in large corpora. Output is fed into a 5,000-MWE lexical database
04/2003 - 12/2004 Finding and Processing Multi-word Lexemes
Designed and tested models originally designed for Dutch MWE identification (open domain) on Hungarian MWES (specific domain)
11/2000 - 05/2005 Algorithms for Linguistic Processing (ALP) PIONIER NWO project

Workshop organization

2006 COLING-ACL Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties
2005 Algorithms for Linguistic Processing PIONIER Workshop

Journal guest editor

2008 Special Issue of the International Journal of Language Resources and Evaluation. "Multiword Expressions: hard going or plain sailing?".

Reviewed for

2008 Int. Conference Lexical Resources and Evaluation (LREC'08)
Workshop "Towards a Shared Task for Multiword Expressions" at Int. Conference Lexical Resources and Evaluation (LREC'08)
2007 45th Annual Meeting of the Association for Computational Linguistics Conference
Conference on Empirical Methods in Natural Language Processing
ACL Workshop on "A broader perspective on Multiword Expressions"
Int. Conference Lexical Resources and Evaluation (LREC)
2006 Int. Journal "Lexical Resources and Evaluation"
Int. Conference Lexical Resources and Evaluation (LREC)
2005 Natural Language Engineering and ESSLI Student Session
2004 Int. Journal of Lexicography
Association for Computational Linguistics Student Workshop

	<i>Teaching</i>	
	Spring 2007	<i>Corpus Linguistics</i> (BA Information Science, 1st year)
	Spring 2006	<i>Corpus Linguistics</i> (BA Information Science, 1st year)
	Spring 2005	<i>Introduction to R statistics package</i> interactive tutorial given for postgraduate students and senior researchers
	Spring 2004	<i>Introduction to Natural Language Processing</i> (BA Information Science, 3rd year) co-taught with Gosse Bouma
	2000 - 2004	<i>Introduction to Natural Language Processing</i> Teaching assistant. Leading computer lab exercises and grading
	1998 - 2000	Trinity College Dublin Lab assistant for <i>Natural Language Processing</i> (BA in Computer Science, Linguistics and a Language, 2nd year)
	1998 - 2000	Lab assistant for <i>Introduction to Computing (Prolog, C, UNIX)</i>
	1997 - 1998	Lab assistant for <i>Artificial Intelligence</i> (BA Computer Science, 3rd year) <i>English language</i> teacher at private school in Santiago de Compostela
HONORS AND AWARDS	1998 - 2000	Trinity College Postgraduate Award
	1993 - 1997	Spanish Ministry of Education Scholarship
	1990 - 1993	Spanish Ministry of Education Scholarship

COMPUTER SKILLS

- Statistical Packages: R, S-Plus.
- Languages: Prolog, Perl, Unix shell scripting (Makefile, awk, sed), XPATH query language.
- Applications: \LaTeX and presentation software.
- Operating Systems: Unix/Linux, Windows.

LANGUAGES

- **Galician** native
- **Spanish** native
- **English** proficient
- **Dutch** intermediate
- **German** beginning

MEMBERSHIPS

Association for Computational Linguistics.

Almere, September 2008.

- Tim van de Cruys and Villada Moirón, B. 2007. Semantics-based Multiword Expression Extraction. In F. van Eynde, P. Dirix, I. Schuurman and V. Vandeghinste (eds.) Proceedings of Computational Linguistics in The Netherlands (CLIN 2006). University of Leuven, 2006.
- Tim van de Cruys and Villada Moirón, B. 2007. Semantics-based Multiword Expression Extraction. In Proceedings of the ACL 2007 Workshop on “A Broader Perspective on Multiword Expressions”, pp. 25-32 . Prague, Czech Republic.
- Villada Moirón, B., A. Villavicencio, D. McCarthy, S. Evert and S. Stevenson (eds.). 2006. Proceedings of the COLING/ACL 2006 Workshop ”Multiword Expressions: Identifying and Exploiting Underlying Properties”. Sydney, Australia.
- Villada Moiron, Begoña and Jörg Tiedemann. 2006. Identifying idiomatic expressions using automatic word-alignment. In Proceedings of the EACL 2006 Workshop on ”Multi-word-expressions in a multilingual context”. Trento, Italy.
- Villada Moirón, Begoña. 2005. Linguistically enriched corpora for establishing variation in support verb constructions. In Proceedings of the 6th International Workshop on Linguistically Interpreted Corpora (Linc’05) held at The 2nd International Joint Conference on Natural Language Processing (IJCNLP-05). Jeju, South Korea.
- Villada Moirón, Begoña. 2005. Data-driven identification of fixed expressions and their modifiability. PhD Thesis. Groningen Dissertations in Linguistics, 52. University of Groningen.
- Kis, Balázs, Begoña Villada, Gosse Bouma, Gábor Ugray, Tamás Bíró, Gábor Pohl and John Nerbonne. 2004. Methods for the extraction of Hungarian Multi-Word Lexemes. In Bart Decadt, Veronique Hoste and Guy de Pauw (eds.) Proc. of Computational Linguistics in the Netherlands 2003. Antwerp Papers in Linguistics. 2004. pp. 47–62
- Villada Moirón, Begoña. 2004. Distinguishing prepositional complements from fixed arguments. In Proc. of the 11th EURALEX International Congress. Lorient, France. pp. 935-942
- Kis, Balázs, Begoña Villada, Gosse Bouma, Gábor Ugray, Tamás Bíró, Gábor Pohl and John Nerbonne. 2004. A New Approach to the Corpus-based Statistical Investigation of Hungarian Multi-word Lexemes. In Proc. of 4th International Conference on Language Resources and Evaluation (LREC 2004). Lisbon, Portugal. pp. 1677–1681.
- Villada Moirón, Begoña. 2004. Discarding noise in an automatically acquired lexicon of support verb constructions. In Proc. of the 4th International Conference on Language Resources and Evaluation (LREC 2004). Lisbon, Portugal. pp. 1859–1862
- Villada, Begoña and Bouma, Gosse. 2002. A corpus-based approach to the acquisition of collocational prepositional phrases. Proc. of EURALEX 2002, Copenhagen. Denmark.
- Bouma, Gosse and Villada, Begoña. 2002. Corpus-based acquisition of collocational prepositional phrases. In Proc. of Computational Linguistics in the Netherlands (CLIN) 2001, University of Twente, 2002.
- Villada, Begoña and Carl Vogel. 2001. Grammatical Relations and Semantic Argument Structure in Spanish Emotion Predicates, in Gutierrez-Rexach, Javier (ed.) Meaning and the Components of Grammar. Significado y los componentes de la Gramática. LINCOS Studies in Theoretical Linguistics 26, Muenchen, Germany:LINCOS Europa.
- Vogel, Carl and Begoña Villada. 2000. Spanish Psychological Predicates. In Ronnie Cann, Claire

Grover, and Philip Miller (eds.) Grammatical Interfaces in HPSG. Stanford, CA:CSLI Publications.

Vogel, Carl and Begoña Villada. 1999. An HPSG analysis of grammatical relations, syntactic valency and semantic argument structure in Spanish psychological predicates and other instances of quirky case and agreement. Tech. Rep. TCD-CS-1999-77; Computational Linguistics Lab. Dept of Computer Science, Trinity College, University of Dublin.

UNPUBLISHED

Villada, Begoña. 2000. Status of Galician and Spanish Pronominal Clitics. MSc dissertation submitted to the University of Dublin.