Curriculum Vitae

First Name: Masood **Nationality:** Iranian
Last Name: Ghayoomi **Marital Status:** Married
Date of Birth: 18.04.1979 **Working Place:** Institute for Humanities and Cultural Studies
Place of Birth: Ghom **Email:** M.Ghayoomi@ihcs.ac.ir

Research Interests

- Natural Language Processing

- Language Modeling

- Machine Learning

- Corpus Linguistics

- Syntax

- Lexical Semantics

Education

2010–2014	**Ph.D. in Computer Science (Computational Linguistics)** Department of Mathematics and Computer Science Freie Universität Berlin, Berlin, Germany *Supervisors*: Prof. Dr. Raúl Rojas, Prof. Dr. Stefan Müller, Prof. Dr. Kiril Ivanov Simov
2012–2013	**Research Visiting** Institute for Natural Language Processing Universität Stuttgart, Stuttgart, Germany *Supervisor*: Prof. Dr. Jonas Kuhn
2010	**Research Visiting** Linguistics Modeling Department Bulgarian Academy of Sciences *Supervisor*: Prof. Dr. Kiril Ivanov Simov
2008–2009	**MSc in Language Science and Technology** Department of Computational Linguistics and Phonetics Universität des Saarlandes, Saarbrücken, Germany *Supervisor*: Prof. Dr. Manfred Pinkal, Dr. Caroline Sporleder
2007–2008	**MSc in Computer, Cognition, and Application** (specialization in Natural Language Processing) Department of Computer Science Nancy2 Université, Nancy, France *Supervisor*: Dr. Bruno Guillaum

2001–2004	**MA in General Linguistics**
	Department of Foreign Languages
	Tehran Azad University, Tehran, Iran
	Supervisors: Prof. Dr. Mostafa Assi, Prof. Dr. Mahmood Bijankhan

1997–2001	**BA in English Language Translation (English-Persian)**
	Department of Foreign Languages
	Ghom Azad University, Ghom, Iran

| 1993–1997 | **Secondary School Diploma in Experimental Sciences** |
| | Khātam-al-anbiā High School, Ghom, Iran |

Awards and Scholarships

| 2018 | **Award of Dr. Kazemi Ashtiyani** |
| | Iran's National Elites Foundation, Tehran, Iran |

| 2012–2013 | **SFB 732 "D2 Combining Contextual Information Sources for Disambiguation in Parsing and Choice in Generation"** |
| | Stuttgart Universität, Stuttgart, Germany |

| 2009–2012 | **DFG for PerGram** |
| | Freie Universität Berlin, Berlin, Germany |

2007–2009	**EU Erasmus Mundus in 'Language and Communication Technologies**
	Nancy2 Université, Nancy, France (2007-2008)
	Universität des Saarlandes, Saarbrücken, Germany (2008-2009)

Working Experience

| 2018–present | **Head of the Secretariat Office for World Universities' Ranking** |

| 2018–present | **Member of the International Office Council** |

2015–present	**Assistant Professor**
	Applied Linguistics Group
	Department of Linguistics
	Institute for Humanities and Cultural Studies, Tehran, Iran
	Project 1: Studying issues for automatic processing of the Persian language on lemmatization, part-of-speech tagging, and parsing: Developing a software using machine learning methods
	Project 2: Unsupervised machine learning methods to induce word senses from a linguistic corpus

2014	**Research Assistant**
	Computer Science Department
	Cooperative State University, Karlsruhe, Germany
	Project: Automated Spelling Diagnosis
	Supervisors: Prof. Dr. Kay Berkling, Junior Prof. Dr. Johanna Fay, Dr. Sebastian Stücker
	Task: In this project, I developed a web-based Graphical User Interface (GUI) to extract data from a linguistics data base and Email it to the user. To develop this GUI, I used HTML, PHP, Javascript, and Perl.

2013–2014	**Research Assistant** Institute for German Language and Literature Karlsruhe University of Education, Karlsruhe, Germany *Project*: Automated Spelling Diagnosis *Supervisors*: Prof. Dr. Kay Berkling, Junior Prof. Dr. Johanna Fay, Dr. Sebastian Stücker *Task*: In this project, I managed and organized a collection of linguistics data to make it possible to extract the required information from the data set. To this end, I wrote scripts by Perl and Javascript to process the data and extract the required information.
2009–2012	**Research Assistant** German Grammar Group Freie Universität Berlin, Berlin, Germany *Project*: Theory and Implementation of an HPSG-grammar for Persian (PerGram) *Supervisors*: Prof. Dr. Stefan Müller, Prof. Dr. Pollet Samvalian *Task*: In this project, I introduced methods in order to extract automatically lexical knowledge and linguistics information from a corpus. Moreover, I introduced methods to reduce human intervention to prepare the data.
2003–2007	**Research Assistant** Institute for Humanities and Cultural Studies, Tehran, Iran *Project*: Expanding the Persian Linguistics Data Base (PLDB) *Supervisor*: Prof. Dr. Mostafa Assi *Task*: In this project, I tried to manage and organize linguistics data. To this end, I developed a data base. Moreover, I was the coordinator of the project.
2005–2007	**Research Assistant** Samt Organization, Tehran, Iran *Project*: Compiling the Persian Advanced Learners' Dictionary *Supervisor*: Prof. Dr. Mostafa Assi *Task*: In this project, I tried to manage and organize linguistics data. I mainly functioned as the coordinator of the project.
2006–2007	**Interpreter and Translator** Master Company, Tehran, Iran *Task*: In the Master Company, an importer of artificial leather, I worked as English-Persian interpreter and translator.
2005	**Corpus Developer** Samt Organization, Tehran, Iran *Project*: Corpus Development from Persian on-line Newspaper Archives *Task*: In this project, I collected a huge amount of linguistics data from Web and organized it to be used for computational linguistics purposes. I also worked as a linguistic consultant.
2005	**Research Assistant** Shegerf Company, Tehran, Iran *Task*: In the Shegerf company, I worked as a linguistic consultant to develop a language resource for the Persian language.

Teaching Experience

2016–now	Institute for Humanities and Cultural Studies, Tehran, Iran Basic Computer
2017	The University of Tehran, Tehran, Iran Knowledge Management and Ontology
2016–2017	Sharif University of Technology, Tehran, Iran Advanced Syntax Morphology
2016	AlZahra University, Tehran, Iran Computational Linguistics (PhD level) Individual Study (PhD level)
21.1.2017 to 25.2.2017	Institute for Humanities and Cultural Studies Practical Workshop on LaTeX
6.11.2016	Institute for Humanities and Cultural Studies Workshop on LaTeXfor Linguists
2016	Islamic Azad University (South Branch), Tehran, Iran: Applied Linguistics (PhD level)
2003–2007	University of Applied Science and Technology, Tehran, Iran (The lectures were in English): Reading Comprehension I, II, and III Conversation I, II Laboratory Grammar I Phonetics General English
2003	Sepid Language School Sadrā Language School
1998	Gooyesh Language School

Invited Talks, Presentations, Demos in Conferences and Workshops

26.12.2017	**Digital Humanities: Application of the Information Technology in the Humanities** Institute for Humanities and Cultural Studies, Tehran, Iran
14.10.2014	**Natural Language Processing and the Persian Language: Requirements and Challenges** IranDoc, Tehran, Iran
17.07.2014 17.12.2012 14.12.2012	**From HPSG-based Persian Treebanking to Parsing: Machine Learning for Data Annotation** Universität des Saarlandes, Saarbrücken Goethe Universität, Frankfurt Universität Stuttgart, Stuttgart

	Introducing the HPSG-based Persian Treebank and its Development Methodology
11.06.2012	Information Technology Research Center, Tehran, Iran
10.06.2012	Institute for Humanities and Cultural Studies, Tehran, Iran
03.06.2012	Sharif University of Technology, Tehran, Iran
	Design and Implementation of Interaction Grammar for Persian
31.12.2009	LSI 1st Seminar on Persian in Cyber-space, Institute for Humanities and Cultural Studies, Tehran, Iran
24-26.06.2008	NaTAL Workshop, Nancy2 Université, Nancy, France
	Frame Assignment with Active Learning
18-19.06.2009	NaTAL Workshop, Nancy2 Université, Nancy, France
	Software Development for Persian Word Prediction
27-28.06.2006	Demonstrated at the 2nd Workshop on Persian and Computer, cooperated with Ehsan Daroodi, The University of Tehran, Tehran, Iran
25-26.05.2004	Demonstrated at the 1st Workshop on Persian and Computer, The University of Tehran, Tehran, Iran

Additional Activities

	Conference Chair
2018	4th National Conference on Computational Linguistics
	Member of the Program Committee, Journal or Conference Reviewer
2017	A Journal of Comparative Linguistic Researchers
	Journal of Language Studies
2016	Farhange Moāser Encyclopedia
2015	Journal of Language and Linguistics
	International Journal of Signal and Data Processing
2013	Behavior Research Methods
2009	Corpora Journal
2018	4th National Conference on Computational Linguistics
2017	9th Conference on Language and Knowledge Technology
2016	2nd National Conference on Corpus Linguistics
2014	3rd Conference of Computational Linguistics, Tehran, Iran
	Scientific Book Reviewer
2017	Samt Press
	Project Supervisor
2016–2017	Developing a Persian Treebank with a Persian Syntactic Parser by the University of Tehran
2016–2017	Developing a Persian Treebank with a Persian Syntactic Parser by the Khaje Nasire Toosi University
	Member of the Executive Committee
11.2007	the 7th Conference of Linguistics at Allāme-Tabātabāyi University
27-28.06.2006	the 2nd Workshop on Persian and Computer at the University of Tehran
14-15.03.2006	the 1st Conference of the Linguistics Society of Iran at the University of Tehran

	Member of Scientific or Student Societies
2017–now	Member of the Computer Society of Iran
2003–now	Member of the Linguistic Society of Iran
2010–2016	Member of the Union of Islamic Students' Associations (in Europe)
2005–2007	Layout formatter of the Journal of Language and Linguistics belonging to the Linguistic Society of Iran
2003–2007	Representative of the Computational Linguistics Committee in LSI

Languages

Persian	Native
English	Fluent
German	Intermediate
French	Beginner

Computer Knowledge

OS	Mac (Very Good), Linux (Very Good), Windows (Very Good)
Programming L.	Perl (Very Good), Python (Good), Java (Good), Prolog (Basic)
Markup L.	HTML (Very Good), XML (Very Good)
Web L.	Javascript (Good), PHP (Good)
Query L.	XPath (Very Good), SPARQL (Basic)
Editors	LaTeX (Very Good), Emacs (Very Good), MS Office (Very Good), LibreOffice (Very Good)
Data Base	Access (Very Good), MySQL (Basic)

Publications

Dissertations

2014	Masood Ghayoomi. *From HPSG-based Persian Treebanking to Parsing: Machine Learning for Data Annotation.* PhD thesis, Freie Universität Berlin, Berlin, Germany, 2014
2009	Masood Ghayoomi. *Frame Assignment with Active Learning.* Master's thesis, Universität des Saarlandes, Saarbrücken, Germany, 2009
2008	Masood Ghayoomi. *Interaction Grammar for the Persian Language.* Master's thesis, Nancy2 Université, Nancy, France, 2008
2004	Masood Ghayoomi. *pišbiniye vāže dar pardāzeše rāyāneyiye zabāne fārsi ["Word Prediction in Computational processing of the Persian language"].* Master's thesis, Islamic Azad University, Tehran Central Branch, Iran, 2004

Papers as Book Chapters or Journals

2018	Saghar Sharifi, Masood Ghayoomi, and Marziye Sana'ati. tanavvo?e negāreši dar fārsi va tahiyeye peykarehāye zabāni dar qālebe farhang ["spelling variation in Persian and developing a corpus in the form of a dictionary"]. *Name-ye Farhangestan: The Special Issue on Lexicography*, (12), forthcoming
	Ahmad Pouramini, Masood Ghayoomi, and Amineh Naseri. Converting dependency treebank to constituency treebank for persian. *Signal and Data Processing*, 14(4):79–96, 2018

2017	Masood Ghayoomi. barresiye mogāyese?iye ta?sire barčasbzaniye maqulāte dasturi bar taǰziye dar pardāzeše xodkāre zabāne fārsi [a comparative study on the impact of part-of-speech tagging on parsing for the Persian language processing]. *Journal of Signal and Data Processing*, 13(4):121–132, 2017
2016	Masood Ghayoomi. tahiyyeye xodkāre zirpeykare is pekareye zabāni [automatic development of a sub-corpus from a language corpus]. In *Proceedings of the 2nd National Conference on Corpus Linguistics*, pages 167–184, Tehran, Iram, September 2016. Nevise-Parsi Press
2015	Johanna Fay, Katrin Hein, and Masood Ghayoomi. Wortfrequenz und Rechtschreibleistung ["Word frequency and correct spelling"]. *Muttersprache*, 2:156–170, 2015
2014	Katrin Hein, Johanna Fay, and Masood Ghayoomi. *'Kleine Wörter' – system-linguistische, sprachdidaktische und empirische Untersuchung einer Wortgruppe mit Sonderstellung innerhalb der diagnostischen Rechtschreibanalyse ["'Small Words' – Linguistic system, language teaching and empirical investigation of a word group with special position within the diagnostic spelling analysis"]*, volume 2 of *Theorie, Empirie und Praxis effektiver Rechtschreibdiagnostik ["Theory, Empiricism and Practice of Effective Spelling Diagnostics"]*, pages 43–62. Stauffenburg Verlag GmbH, Tübingen, Deutschland, 2014
2012	Masood Ghayoomi. Bootstrapping the development of an HPSG-based treebank for Persian. *Linguistic Issues in Language Technology*, 7(1), 2012
2011	Mahmood Bijankhan, Javad Sheykhzadegan, Mohammad Bahrani, and Masood Ghayoomi. Lessons from building a Persian written corpus: Peykare. *Language Resources and Evaluation*, 45(2):143–164, 2011
2010	Masood Ghayoomi, Saeedeh Momtazi, and Mahmood Bijankhan. A study of corpus development for Persian. *International Journal on Asian Language Processing*, 20(1):17–33, 2010

Papers in Conferences and Workshops

2018	Masood Ghayoomi. mas?aleye čandvāžegi dar pardāzeše nahve rāyānešiye zabāne fārsi [the problem of multi-words in syntactic processing of Persian]. In *Proceedings of the Fourth Nation Conference on Computational Linguistics*, pages 11–40, Tehran, Iran, February 2018. Neviseh Parsi Press
2016	Masood Ghayoomi. tahiyyeye xodkāre zirpeykare is pekareye zabāni [automatic development of a sub-corpus from a language corpus]. In *Proceedings of the 2nd National Conference on Corpus Linguistics*, pages 167–184, Tehran, Iram, September 2016. Nevise-Parsi Press

2015   Masood Ghayoomi, Saghar Sharifi, and Marziyeh Sanaati. tanavvo?e negāreši dar zabāne fārsi va tahiyyeye xodkāre dādegāne ?emlāyi ?az peykareye zabāniye mobtani bar veb [spelling variation in the Persian language and automatic development of a spelling database from a Web-based corpus]. In *Proceedings of the 1st International Conference on Web Research*, University of Science and Culture, Tehran, Iran, 2015

Saghar Sharifi, Marziyeh Sanaati, and Masood Ghayoomi. Spelling variation problem in Persian: Automatic development of a spelling database for a digital dictionary. In *Proceedings of the 9th International Conference of ASIALEX 2015 Hong Kong: Words, Dictionaries and Corpora : Innovations in Reference Science*, pages 488–497, The Hong Kong Polytechnic University, Hong Kong, 2015

2014   Saeedeh Momtazi and Masood Ghayoomi. Weakly supervised text categorization using topic modeling. In *Proceedings of the 3rd Conference on Computational Linguistics*, Tehran, Iran, 2014

MohammadBagher Sadeghzadeh, MohammadReza Razzazi, and Masood Ghayoomi. Studying the effecting methods on the performance of a statistical parser for the Persian language. In *Proceedings of the 3rd Conference on Computational Linguistics*, Tehran, Iran, 2014

Masood Ghayoomi and Jonas Kuhn. Converting an HPSG-based treebank into its parallel dependency-based treebank. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*, pages 802–809, Reykjavik, Iceland, 2014

Masood Ghayoomi, Kiril Simov, and Petya Osenova. Constituency parsing of Bulgarian: Word- vs class-based parsing. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*, pages 4056–4060, Reykjavik, Iceland, 2014

Kay Berkling, Johanna Fay, Masood Ghayoomi, Katrin Hein, Rémi Lavalley, Ludwig Linhuber, and Sebastian Stüker. A database of freely written texts of German school students for the purpose of automatic spelling error classification. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*, pages 1212–1217, Reykjavik, Iceland, 2014

Johanna Fay, Kay Berkling, Sebastian Stücker, Masood Ghayoomi, Katrin Hein, and Rémi Lavalley. Zwei Perspektiven auf Rechtschreibung: Sprachdidaktik und Computerlinguistik entwickeln ein Instrument zur Rechtschreibdiagnostik ["Two perspectives on correct spelling: Using speech recognition and computational linguistics to develop an instrument for spelling diagnostics"]. In *Zweiten Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF)*, Frankfurt am Main, Germany, 2014

2013   Masood Ghayoomi and Jonas Kuhn. Sampling methods in active learning for treebanking. In *Proceedings of the 12th International Workshop on Treebanks and Linguistic Theories*, pages 49–60, Sofia, Bulgaria, 2013

Masood Ghayoomi. Introducing a treebank and a statistical parser for Persian. In *Proceedings of the 8th Conference of Iranian Linguistics*, volume 2, pages 666–679, 2013

2012	Masood Ghayoomi and Omid Moradiannasab. The effect of treebank annotation granularity on parsing: A comparative study. In *Proceedings of the 11th International Workshop on Treebanks and Linguistic Theories*, pages 109–114, Lisbon, Portugal, 2012. Edições Colibri
	Masood Ghayoomi. Word clustering for Persian statistical parsing. In Hitoshi Isahara and Kyoko Kanzaki, editors, *Advances in Natural Language Processing*, volume 7614 of *Lecture Notes in Computer Science: JapTAL '12: Proceedings of the 8th International Conference on Advances in Natural Language Processing*, pages 126–137. Springer Berlin Heidelberg, 2012
	Masood Ghayoomi. From grammar rule extraction to treebanking: A bootstrapping approach. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pages 1912–1919, Istanbul, Turkey, 2012
2011	Masood Ghayoomi and Stefan Müller. Multi-token units and multi-unit tokens in developing an HPSG-based treebank for Persian. In *4th International Conference on Iranian Linguistics*, page 29, 2011
2010	Stefan Müller and Masood Ghayoomi. PerGram: A TRALE implementation of an HPSG fragment of Persian. In *Proceedings of 2010 IEEE International Multi-conference on Computer Science and Information Technology*, pages 461–467, Wisła, Poland, 2010
	Masood Ghayoomi. Using variance as a stopping criterion for active learning of frame assignment. In *Proceedings of the Human Language Technology-Conference of the North American Chapter of the Association for Computational Linguistics, Workshop on Active Learning for Natural Language Processing*, pages 1–9, Los Angeles, USA, 2010
2009	Masood Ghayoomi and Saeedeh Momtazi. Challenges in developing Persian corpora from online resources. In *Proceedings of 2009 IEEE International Conference on Asian Language Processing*, pages 108–113, 2009
	Masood Ghayoomi and Saeedeh Momtazi. An overview on the existing language models for prediction systems as writing assistant tools. In *Proceedings of 2009 IEEE International Conference on Systems, Man, and Cybernetics*, pages 5233–5237, 2009
	Masood Ghayoomi and Bruno Guillaume. Interaction grammar for the Persian language: Noun and adjectival phrases. In *Proceedings of the Association for Computational Linguistics and International Joint Conference on Natural Language Processing, 7th Workshop on Asian Language Resources*, pages 107–114, Suntec, Singapore, 2009. Association for Computational Linguistics
2008	Masood Ghayoomi and Ehsan Daroodi. A POS-based word prediction system for the Persian language. In Bengt Nordström and Aarne Ranta, editors, *Advances in Natural Language Processing*, volume 5221 of *Lecture Notes in Computer Science: GoTAL '08: Proceedings of the 6th International Conference on Advances in Natural Language Processing*, pages 138–147. Springer Berlin Heidelberg, 2008
2007	Shahram Modarres-Khiabani and Masood Ghayoomi. The role of statistics as means for studying lexical relations. In *Proceedings of the 7th Conference of Iranian Linguistics*, pages 604–618, 2007

2006	Masood Ghayoomi. Using word prediction systems for users with disabilities: A case study. In *Proceedings of the 2nd Workshop on the Persian Language and Computer*, pages 216–225, 2006
	Niloufar Mansoori-Haredasht, Masood Ghayoomi, and Zahra Abolhasani-Chimeh. Using conceptual graphs for semantic representation of Persian prepositions. In *Proceedings of the 1st Conference of the Linguistics Society of Iran*, pages 105–115, 2006
	Zahra AbolhasaniChime and Masood Ghayoomi. Incorporation: Word production of Persian prepositions and its application in computational linguistics. In *Proceedings of the 2nd Workshop on the Persian Language and Computer*, pages 16–24, 2006
	Shahram Modarres-Khiabani and Masood Ghayoomi. Using language corpora in studying collocation: A comparative approach. In *Proceedings of the 2nd Workshop on the Persian Language and Computer*, pages 55–65, 2006
2005	Masood Ghayoomiand SeyyedMostafa Assi. Word prediction in a running text: A statistical language modeling for the Persian language. In *Proceedings of the Australasian Language Technology Workshop*, pages 57–63, Sydney, Australia, 2005
2004	Masood Ghayoomi. *pišbiniye vāže dar pardāzeše rāyāneyiye zabāne fārsi ["Word Prediction in Computational processing of the Persian language"]*. Master's thesis, Islamic Azad University, Tehran Central Branch, Iran, 2004

Project Reports

2009	Hole Semantics and Minimal Recursion Semantics: A Comparison Approach Underspecification Seminar Universität des Saarlandes, Saarbrücken, Germany
2008	TAG for the Persian Language Linguistics Course Nancy2 Université, Nancy, France
	XSL for Transforming XML Recognition Results Web Technology Course Nancy2 Université, Nancy, France
2007	XMG Grammar for English NLP Tools and Application Course Nancy2 Université, Nancy, France

April 28, 2018
Tehran, Iran
Masood Ghayoomi