

Şaziye Betül Özates

CONTACT INFORMATION

Institute for Data Science and AI
Boğaziçi University
Güney Kampüs, Bebek
34342 İstanbul, Turkey
saziye.ozates@bogazici.edu.tr
[linkedin.com/in/saziye-betül-özates-a86982211](https://www.linkedin.com/in/saziye-betül-özates-a86982211)

RESEARCH INTERESTS

Natural language processing, deep learning, machine learning, computational linguistics, artificial intelligence.

ABOUT

I am a computer scientist who specializes in natural language processing and machine learning. I have received my bachelor's degree, and master's degree from the Department of Computer Engineering at Boğaziçi University. I completed my PhD studies in the same place under the guidance of Dr. Arzucan Özgür and Dr. Tunga Güngör. From 2020 to 2022, I was a researcher at the Institute of Natural Language Processing at the University of Stuttgart and worked as a post-doctoral research fellow at KUIS-AI Center between 2022 and 2023. Currently, I am an assistant professor at the Institute for Data Science and Artificial Intelligence at Boğaziçi University.

EXPERIENCE

Visiting Scholar

IMS, University of Stuttgart, 2025-2025

Assistant Professor - Full Time

DSAI, Boğaziçi University, 2023-

Research Fellow - Full Time

KUIS AI Center, 2022-2023

Researcher - Full Time

IMS, University of Stuttgart, 2020-2022

Teaching Assistant - Part Time

Boğaziçi University, 2015-2016, 2020

Research Assistant - Full Time

Turkish-German University, 2016-2018

EDUCATION

Boğaziçi University, İstanbul, Turkey

Ph.D. in Computer Engineering, May 2022

- Thesis Topic: Deep Learning-based Dependency Parsing for Turkish
- Advisor: Assoc. Prof. Arzucan Özgür and Prof. Tunga Güngör
- Earned 4.0 GPA on 4.0 scale in coursework

Boğaziçi University, İstanbul, Turkey

M.S. in Computer Engineering, June 2014

- Thesis Topic: Multi-document Summarization using Dependency Grammars
- Advisor: Assoc. Prof. Arzucan Özgür
- Earned 4.0 GPA on 4.0 scale in coursework

Boğaziçi University, İstanbul, Turkey

B.S. in Computer Engineering, June 2012

- Senior Project Topic: A Comparison of Financial Time Series Forecasting Using Multilayer Perceptron, Support Vector Machines, and Bayesian Network
- Advisor: Prof. Fikret Gürgen
- Highest honors in Engineering Faculty
- Ranked 2nd in graduating class with 3.64 GPA on 4.0 scale

PUBLICATIONS

Refereed International Journal Papers

Atam, E., Zeybek, A., Özateş, Ş. B., Güzelkaya, N. (2026). "A Deep Learning Approach for Gain-Scheduled PID Control Design via Advanced Supervisory Controllers", *IEEE Transactions on Emerging Topics in Computational Intelligence*, *Accepted*.

Özateş, Ş. B., Özgür, A., Güngör, T., Öztürk, B. (2022). "A Hybrid Approach to Dependency Parsing: Combining Rules and Morphology with Deep Learning." *IEEE Access*, *10*, 93867-93886.

Türk, U., Atmaca, F., Özateş, Ş. B. *et al.* (2022). "Resources for Turkish Dependency Parsing: Introducing the BOUN Treebank and the BoAT Annotation Tool." *Language Resources & Evaluation*, *56(1)*, 259-307.

Selected Refereed International Conference Papers

Öztürk, M. A., Özateş, Ş. B. Root, S. B., *et al.* (2026). "Automated Screening of Antibacterial Nanoparticle Literature: Dataset Curation and Model Evaluation." *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2026)*. Association for Computational Linguistics.

Bilgin Taşdemir, E. F., Gürer, D. Z., Özateş, Ş. B. (2026). "RuznamceNER: A Named Entity Recognition Dataset for Ottoman Turkish." *Proceedings of the Fifteenth International Conference on Language Resources and Evaluation (LREC 2026)*. European Language Resources Association (ELRA).

Tıraş, T. E., Ünal N. C., Cengiz, A., Yurtseven, E., Bilgin Taşdemir, E. F., Özateş, Ş. B. (2026). "OTA-BOUN: A Historical Turkish Dependency Treebank." *Proceedings of*

the Fifteenth International Conference on Language Resources and Evaluation (LREC 2026). European Language Resources Association (ELRA).

Bilgin Taşdemir, E. F., Özateş Ş. B. (2025). "NakbaTR: A Turkish NER Dataset for Nakba Narratives." *Proceedings of the first International Workshop on Nakba Narratives as Language Resources, pages 122–126, Abu Dhabi. Association for Computational Linguistics.*

Özateş, Ş. B., Tıraş, T. E., Genç, E. F., Bilgin Taşdemir, E. F. (2024). "Dependency Annotation of Ottoman Turkish with Multilingual BERT." *Proceedings of the 18th Linguistic Annotation Workshop, St. Julians, Malta. Association for Computational Linguistics.*

Özateş, Ş. B., Özgür, A., Güngör, T., Çetinoğlu, Ö. (2022). "Improving Code-Switching Dependency Parsing with Semi-Supervised Auxiliary Tasks." *Findings of the Association for Computational Linguistics: NAACL 2022, pp. 1159-1171.*

Özateş, Ş. B., Çetinoğlu, Ö. (2021). "A Language-aware Approach to Code-switched Morphological Tagging." *Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching, pp. 72-83. Association for Computational Linguistics.*

Türk, U., Atmaca, F., Özateş, Ş. B. *et al.* (2019). "Improving the Annotations in the Turkish Universal Dependency Treebank." *In Universal Dependencies Workshop (UDW 2017) at SyntaxFest 2019, pp. 108-115, Paris, France. Association for Computational Linguistics.*

Türk, U., Atmaca, F., Özateş, Ş. B. *et al.* (2019). "Turkish Treebanking: Unifying and Constructing Efforts." *Proceedings of the 13th Linguistic Annotation Workshop, pp. 166-177, Florence, Italy. Association for Computational Linguistics.*

Özateş, Ş. B., Özgür, A., Güngör, T., Öztürk, B. (2018). "A Morphology-based Representation Model for LSTM-based Dependency Parsing of Agglutinative Languages." *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, pp. 238-247, Brussels, Belgium. Association for Computational Linguistics.*

Özateş, Ş. B., Özgür, A., Güngör, Radev, D. R. (2016). "Sentence Similarity based on Dependency Tree Kernels for Multi-document Summarization." *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), pp. 2833-2838, Portorož, Slovenia. European Language Resources Association.*

Please see the full publication list from <https://scholar.google.com/citations?user=EAlmj9yYJPOC&hl=en&oi=ao>

Thesis

Şaziye Betül Özateş. "Deep Learning-based Dependency Parsing for Turkish", *Ph.D. Thesis, Computer Engineering, Boğaziçi University, İstanbul, Turkey, May 2022.*
Thesis Advisors: Assoc. Prof. Arzucan Özgür and Prof. Tunga Güngör.

Şaziye Betül Bilgin. "Multi-document Summarization using Dependency Grammars", *M.S. Thesis, Computer Engineering, Boğaziçi University, İstanbul, Turkey, June 2014.*
Thesis Advisor: Assoc. Prof. Arzucan Özgür.

BAP PROJECT

- Deep Learning-based Exploration of Linguistic Structures and Semantic Entities in Historical Turkish Texts (2024 -)
 - This project aims to analyze the structural and semantic features of historical Turkish documents from the Ottoman Empire and early Republic periods using deep learning. In the scope of this project, various datasets were manually annotated to address the lack of labeled data. These resources support the training and evaluation of transformer-based models for detailed syntactic and semantic analysis. All annotated data and models have been made publicly available at <https://huggingface.co/BUCOLIN>, supporting research in computational linguistics and historical studies.
Role in the project: Project Manager

DFG PROJECT (GERMAN RESEARCH FOUNDATION)

- Computational Structural Analysis of German-Turkish Code-Switching (2020 - 2022)
 - The aim of this project is to analyse Turkish-German code-switching (CS) from a computational perspective. We developed deep learning-based computational models for morphological and syntactic analyses of CS data.
Role in the project: Researcher

TÜBİTAK PROJECTS

- ARDEB 3005 - Analysis of Calligraphy Artworks Using Artificial Intelligence: Understanding Cultural Heritage and Accessibility 2026 -
 - This project aims to digitize Arabic-script calligraphy inscriptions found in historical architectural structures in Türkiye, automatically recognize their content using artificial intelligence, and interpret them through multilingual translation. Through an AI-powered mobile application, users will be able to obtain instant information from inscription images they capture with their mobile phones; at the same time, the database will be enriched through community contributions. Within the project, advanced AI methods such as CBIR, open-set classification, VQA, and LLMs will be employed to create a novel database of artistic inscription images, contributing —as a first in Türkiye— to the digitization and sustainability of cultural heritage.
Role in the project: Researcher
- ARDEB 3005 - Automatic Processing and Analysis of Kazasker Ruznamçe Records with Digital Methods (2024 -)
 - This project aims to extract information from the Anatolian Kazasker Ruznamçe Registers covering the years 1660–1895 using automated and semi-automated methods. The extracted data will then be linked and utilized to analyze the Ottoman state’s economic, administrative, and institutional structures during this period. For the first time, this will enable a continuous examination of approximately 250 years. Additionally, the project plans to semantically model the collected data, transfer it into a graph-based database, discover new knowledge within the data, and integrate it into Linked Open Data sources, offering an innovative perspective to the field.
Role in the project: Researcher
- BİLGEM - Development of a Deep Learning-Based Data Expansion Tool for Predicting the Efficacy of Antibacterial Nanoparticles Used in Nanomedicine (2024 - 2025)

- This project aims to build a comprehensive AI-based database to systematically map nanoparticle (NP) and bacteria interactions, helping to develop machine learning models that predict NP effectiveness. The project will also offer a standardized checklist for antibacterial NP synthesis, facilitating faster research progression into clinical studies. Ultimately, the database aims to uncover the mechanisms governing NP-bacteria interactions and accelerate the development of new antimicrobial NP agents against infectious diseases. *Role in the project: Advisor*
- ARDEB 1005 - A Deep Learning based Turkish Dependency Parser (2018 - 2019)
 - We developed a novel state-of-the-art deep learning-based dependency parsing system and created a manually annotated dependency treebank for Turkish. With this system and the newly introduced treebank, we reached the best performance on the dependency parsing task that is needed in many natural language processing applications using deep learning techniques for Turkish.
Role in the project: Main Researcher

INTERNATIONAL CONFERENCES

- ACL, The 63rd Annual Meeting of the Association for Computational Linguistics, July 31 – August 1, 2025, Vienna, Austria.
- COLING, The 31st International Conference on Computational Linguistics, January 19–24, 2025, Online.
- EACL, The 18th Conference of the European Chapter of the Association for Computational Linguistics, March 17-22, 2024, St. Julians, Malta.
- NAACL, 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics, June 6–11, 2021, Online.
- ACL, The 57th Annual Meeting of the Association for Computational Linguistics (ACL) July 28-August 2, 2019, Florence, Italy
- EMNLP, 2018 Conference on Empirical Methods in Natural Language Processing, October 31–November 4, 2018, Brussels, Belgium
- LREC, The 10th edition of the Language Resources and Evaluation Conference, May 23-28, 2016, Portorož, Slovenia

INVITED TALKS

- Şaziye Betül Özateş, "Challenges in Developing Natural Language Processing Resources and Methods for Ottoman Turkish", Translating Ottomans: NLP Contributions to Historical Research on Translingual Practices Workshop, Ca' Foscari University, Venice, Italy (5 Dec, 2025)
- Şaziye Betül Özateş, "Osmanlıca Dil Modellerinin İnşası", Osmanlı ve Türkiye Çalışmalarında Doğal Dil İşleme ve Büyük Dil Modelleri Çalıştayı, İYTE, İzmir, Turkey, (10 Oct, 2025)

- Şaziye Betül Özateş, “NLP for Low-resource Languages using Deep Learning and Linguistic Features”, Graduate Seminar, Boğaziçi University, Istanbul, Turkey (24 Dec, 2024)
- Şaziye Betül Özateş, “A Semi-supervised Approach for Improving Code-switching Dependency Parsing”, Graduate Seminar, Boğaziçi University, Istanbul, Turkey (29 Nov, 2022)
- Şaziye Betül Özateş, “A Semi-supervised Dependency Parsing Approach for Low-resource Languages”, CS and AI Talks, Boğaziçi University, Istanbul, Turkey (11 Nov, 2022)
- Şaziye Betül Özateş, “Deep Learning-based Dependency Parsing For Turkish”, ClingDing Talks, Indiana University, Bloomington, IN, USA (20 Oct, 2022)
- Şaziye Betül Özateş, "Natural Language Processing for Low-resource Languages using Deep Learning and Linguistic Features", KUIS AI Talks, Koç University, İstanbul, Turkey (18 Jul, 2022)
- Şaziye Betül Özateş, "Improving Dependency Parsing of Turkish: A Hybrid Neural Parser and the BOUN Treebank", Institutsversammlung, IMS, Universität Stuttgart, Stuttgart, Germany (4 Nov, 2020)

TEACHING
EXPERIENCE

Instructor

Institute for Data Science and Artificial Intelligence, Boğaziçi University

DSAI 545, Natural Language Processing

DSAI 514, Statistical Inference

DSAI 526, Web Mining

DSAI 303, Probability and Statistics for Data Science and Artificial Intelligence

Teaching Assistant

Computer Engineering Department, Boğaziçi University

CMPE 150, Introduction to Computing

CMPE 493, Introduction to Information Retrieval

HONORS AND
AWARDS

- 2026 PRESTIJ 2026 Research Fellowships Program - Laureate
- 2025 NAACL SRW 2025 - Best Paper Award
- 2018-2019 Tubitak ARDEB 1005 - Scholarship from National New Ideas and Products Research Support Program
- 2014-2019 Tubitak BIDEB 2211 - Scholarship for PhD Students
- 2012-2014 Tubitak BIDEB 2228 - MSc and PhD Scholarship for Senior Undergraduate Students
- 2012 Department Second Rank Award (GPA:3.64/4.00), Computer Engineering, Boğaziçi University
- 2012 Highest honors in Engineering Faculty, Boğaziçi University

EDITORIAL AND
REVIEWING
ACTIVITIES

- Area Chair, LREC 2026, Corpora, Treebanks and Annotation; Tools, Systems and Platforms Track
- Reviewer, ACL ARR, (2023 - present)
- Reviewer, Springer Nature Journals, (2022 - present)
- Reviewer, Elsevier ScienceDirect Journals, (2024 - present)
- Reviewer, IEEE SIU, (2019 - 2024)

COMPUTER
SKILLS

- Python, Java, C, C++
- Matlab, Latex, Perl
- Deep neural networks, Transformers, NLP applications
- OS: Linux-based systems, Windows

RELEVANT
SKILLS

Languages: Turkish (*Native*), English (*Advanced*), German (*Intermediate*)