

EDUCATION**Tel Aviv University**, Tel Aviv, Israel

2007–2011: Adi Lautman Interdisciplinary Program for Outstanding Students scholar.

2011: *Bachelor of Science*, Mathematics and Computer Science (Double Major), *magna cum laude*.2014: *Master of Arts*, Linguistics, *summa cum laude*.Thesis topic: Semantics of *Before*

Advisor: Dr. Nirit Kadmon

Georgia Institute of Technology, Atlanta, GA2016–2021: *Doctor of Philosophy*, Computer Science

School of Interactive Computing

Advisor: Prof. Jacob Eisenstein

Thesis title: *Integrating Distributional, Compositional, and Relational Approaches to Neural Word Representations***AWARDS****Honors**

- 2024: Outstanding Paper Award at IEEE HPEC (for *MonoCoder: Domain-Specific Code Language Model for HPC Codes and Tasks*).
- 2024: Outstanding Paper Award & Senior Area Chair Award at ACL (for *Greed is All You Need: An Evaluation of Tokenizer Inference Methods*).
- 2008–2011: Annual recipient of an Adi Lautman scholarship for excellence (in addition to tuition restitution).
- 2010: Pozis Foundation grant for distinguished MA students.

Fellowships

- 2018, 2019, 2020: Bloomberg Data Science Ph.D Fellowships for 2018–21 (full tuition & full annual stipend & Travel grant; ~ \$77,000 annually).

EMPLOYMENT HISTORY**Senior Lecturer**

September 2021–Present

Ben-Gurion University of the Negev

Be'er Sheva, Israel

PI of the MeLeL lab (Machines Learning Language) at the Faculty of Computer and Information Science, Institute of Applied AI Research.

*Until 2025: Department of Computer Science***Visiting Academic**

January 2023–March 2025

Amazon Science

Tel Aviv, Israel

Working with the Personalization research team at 20% capacity.

Teaching Assistant

August–December 2020 (Head TA)

Georgia Institute of Technology

Atlanta, GA

January–April 2018

CS 4650/7650—Natural Language (Profs. Diyi Yang, Jacob Eisenstein)

Research Science Intern

May–July 2020
May–August 2019

Bloomberg LP

New York, NY

Research Science Intern

May–August 2018

ASAPP, Inc.

New York, NY

Research Intern

May–August 2017

The Home Depot

Atlanta, GA

Research Engineer

April 2013–July 2016

Yahoo Labs / Yahoo Research

Haifa, Israel

Computational Linguist

May 2011–February 2013

Ginger Software

Tel Aviv, Israel

Teaching Assistant

March 2009–September 2010

Tel Aviv University Linguistics

Tel Aviv, Israel

Recitation instructor for introductory and advanced courses in semantics for Dr. Nirit Kadmon.

**SCIENTIFIC
PUBLICATIONS**

Refereed Articles in Conference Proceedings (Selected)

1. David Carmel, Avihai Mejer, **Yuval Pinter**, and Idan Szpektor. *Improving Term Weighting for Community Question Answering using Syntactic Analysis*. Proceedings of the 23rd ACM international conference on conference on information and knowledge management (CIKM), November 2014, pp. 351–360. Shanghai, China.
2. **Yuval Pinter**, Roi Reichart, and Idan Szpektor. *Syntactic Parsing of Web Queries with Question Intent*. Proceedings of the 2016 conference of the North American chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL), June 2016, Issue 1: Long Papers, pp. 670–680. San Diego, CA.
3. **Yuval Pinter**, Robert Guthrie, and Jacob Eisenstein. *Mimicking Word Embeddings using Subword RNNs*. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP), September 2017, pp. 102–112. Copenhagen, Denmark.
4. Sarah Wiegrefe[†] and **Yuval Pinter**[†]. *Attention is not not Explanation*. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), November 2019, pp. 11–20. Hong Kong.
5. Shaked Yehezkel and **Yuval Pinter**. *Incorporating Context into Subword Vocabularies*. Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL): Main Volume. May 2023, pp. 623–635.
6. Tal Kadosh, Niranjan Hasabnis, Timothy Mattson, **Yuval Pinter**, and Gal Oren. *Quantifying OpenMP: Statistical Insights into Usage and Adoption*. IEEE High Performance Extreme Computing. September 2023, pp. 1–7.

7. Lisa Beinborn and **Yuval Pinter**. *Analyzing Cognitive Plausibility of Sub-word Tokenization*. EMNLP, December 2023, pp. 4478–4486.
8. Stephen Mayhew, Terra Blevins, Shuheng Liu, Marek Suppa, Hila Gonen, Joseph Marvin Imperial, Börje F. Karlsson, Peiqin Lin, Nikola Ljubešić, Lester James Validad Miranda, Barbara Plank, Arij Riabi, **Yuval Pinter**. *Universal NER: A Gold-Standard Multilingual Named Entity Recognition Benchmark*. NAACL, June 2024, pp. 4322–4337.
9. Omri Uzan, Craig W. Schmidt, Chris Tanner, **Yuval Pinter**. *Greed is All You Need: An Evaluation of Tokenizer Inference Methods*. ACL, August 2024, pp. 813–822. **Outstanding Paper Award**.
10. Craig W. Schmidt, Varshini Reddy, Haoran Zhang, Alec Alameddine, Omri Uzan, **Yuval Pinter**, Chris Tanner. *Tokenization is More Than Compression*. EMNLP, November 2024, pp. 678–702.

Refereed Articles in Scientific Journals (Selected)

1. Carmel Vaisman, Illan Gonen, and **Yuval Pinter**. *Nonhuman Language Agents in Online Collaborative Communities: Comparing Hebrew Wikipedia and Facebook Translations*. Discourse, Context & Media . March 2018, pp. 10–17.
2. Re’em Harel, Tal Kadosh, Niranjan Hasabnis, Timothy Mattson, **Yuval Pinter**, and Gal Oren. *PragFormer: Data-Driven Parallel Source Code Classification with Transformers*. International Journal of Parallel Programming . February 2025, pp. 1–26.

Patents (Selected)

1. Dan Pelleg, Oleg Rokhlenko, Idan Szpektor, **Yuval Pinter**, David Carmel, Somesh Jain, Archit Shrivastava, Shirin Oskooi. Quality-based scoring and inhibiting of user-generated content. US Patent #10489712, published November 26, 2019.

INVITED LECTURES

Keynote Talks

1. March 2025: *Building Character: Tokenization in Theory and in Practice*. HumanCLAIM Workshop. Göttingen, Germany.
2. July 2025: *Beat them? Join them? Fix them? Tokenization Research in a Downstream World*. Tokenization Workshop at ICML. Vancouver, Canada.

Invited Talks

1. March 2025: *Shukkakhten: Language Models and their Attempts to Understand New Words*. Hebrew Language Academy Conference on Lexicology in Transition. Jerusalem, Israel.
2. July 2025: *Building Character: Tokenization in Theory and in Practice*. Frontiers in NLP. University of British Columbia.

GRANTS

Competitive Grants

- 2025: ISF–DFG Lead Agency Research Grant (for research on *Hyperbolic Interlingual Representations*, with Lucie Flek (Universität Bonn)) (ILS 579,000 ISF side)

- 2023: ISF Personal Grant (for research on *Meaningful Subword Text Tokenization*) (ILS 844,000)
- 2023: BSF Start-up Grant (for research on *Diacritization for the World's Scripts*, with Kyle Gorman (CUNY)) (\$150,000)
- 2022: Ministry of Science Grant (for research on *Privacy Preserving Text Analysis*, with Niv Gilboa (BGU)) (ILS 300,000)

Non-Competitive and Internal Grants

- 2023: BGU VP R&D and Faculty of Humanities and Social Sciences Interdisciplinary Grant (for research on *Finding Clues for Alzheimer's Disease in Early Online Language*, with Talya Sadeh (BGU)) (ILS 25,000)
- 2022: Unrestricted Award, Google Inc. (for research on *Meaningful Subword Text Tokenization*) (\$30,000)
- 2022: BGU Data Science Research Center Grant (for research on *Code Language Parallel Processing*, with Gal Oren (Technion)) (ILS 88,000)
- 2022: Foundation For Higher Education and Culture (FFHEC) Recruitment Award for Outstanding New Scientists (\$100,000)

EDUCATIONAL ACTIVITIES

Courses Taught

2025/B	202-1-2051 Principles of Programming Languages—Undergrad
2024/B	mandatory (4 credits), Ben-Gurion University
2025/A	
2024/A	202-2-5211 Natural Language Processing—Graduate / Undergrad
2023/A	elective (4 credits), BGU
2022/A	
2023/Y	142-1-0172 Foundations of Algorithmic Trading (as CS
2022/B	Coordinator)—Interdisciplinary elective (3–9 credits), BGU

Courses Taught, Non-Academic

2025/March	AI Course for Executives (NLP section), The Institute
------------	---

Research Students, graduated

1. Nadav Schneider, MSc. (joint with Gal Oren, Technion and Dan Vilenchik, BGU EE) (graduated 2025)
2. Bar Gazit, MSc. (2025)
3. Gal Astrach, MSc. (2025)
4. Tal Kadosh, MSc. (joint with Gal Oren) (2025)
5. Carinne Cherf, MSc. (2024)
6. Peter Kogan, MSc. (2024)
7. Shaked Yehezkel, MSc. (joint with Omer Levy, TAU) (2022)

Research Students, current

-
1. Shira Darchi, Ph.D (joint with Talya Sadeh) (candidate since 2024; expected graduation 2027)
 2. Re'em Harel, Ph.D (candidate since 2024; exp. 2028)
 3. Benzi Busigin, MSc. (exp. 2027)
 4. Ibraheem Abo Shakra, MSc. (exp. 2026)
 5. Adi Cohen, MSc. (exp. 2026)
 6. Rotem Brilliant, MSc. (exp. 2026)
 7. Bar Block, MSc. (exp. 2026)
 8. Or Kadosh, MSc. (exp. 2026)
 9. Omri Uzan, MSc. (exp. 2025)
 10. Yair Elboher, MSc. (exp. 2025)
 11. Tamar Levy, MSc. (joint with Jonathan Berant, TAU) (exp. 2025)

LANGUAGES**Natural**

Hebrew: native speaker
 English: native speaker level
 Spanish: conversational
 French, Czech, Russian: beginner

Computer

Java, Python, C#: fluent
 C++, C, PHP, TypeScript, Scheme:
 conversational
 Environments: Windows, Unix

INTERESTS**Professional**

Natural Language Processing: Semantics; Morphology; Interpretability; Syntactic Analysis; User-Generated Content.
 Other CS fields: Machine Learning; Information Retrieval; Data Mining; Algorithms.

Personal

Blogging: 2009 – Present: Contributor, editor at Dagesh Kal, Israeli linguablog (in Hebrew). <http://dagesh.wordpress.com/English>
 Long-distance running: full marathons, 2013 and 2015.
 Trivia: four-time game show contestant (Israeli national TV).