

**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 ACL (for *Greedy 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****Visiting Academic**

January 2023 – Present

Working with the Personalization research team at 20% capacity.

**Amazon Science**

Tel Aviv, Israel

**Senior Lecturer**

September 2021 – Present

PI of the MeLeL lab (Machines Learning Language) at the Department of Computer Science; member, Data Science Research Center.

**Ben-Gurion University of the Negev**

Be'er Sheva, Israel

**Teaching Assistant**

August – December 2020 (Head TA)

January – April 2018

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

**Georgia Institute of Technology**

Atlanta, GA

**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

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. Gilad Tsur, **Yuval Pinter**, Idan Szpektor, and David Carmel. *Identifying Web Queries with Question Intent*. Proceedings of the 25th International Conference on World Wide Web (WWW), April 2016, pp. 783–793. Montreal, Canada.
3. **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.
4. **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.
5. Ian Stewart<sup>†</sup>, **Yuval Pinter**<sup>†</sup>, and Jacob Eisenstein. *Sí o no, ¿què penses? Catalanian Independence and Linguistic Identity on Social Media*. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, June 2018, Volume 2 (Short Papers), pp. 136–141. New Orleans, LA.
6. **Yuval Pinter** and Jacob Eisenstein. *Predicting Semantic Relations using Global Graph Properties*. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, October 2018, pp. 1741–1751. Brussels, Belgium.
7. Sarah Wiegrefe<sup>†</sup> and **Yuval Pinter**<sup>†</sup>. *Attention is not not Explanation*. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natu-

- ral Language Processing (EMNLP-IJCNLP), November 2019, pp. 11–20. Hong Kong.
8. Arya D. McCarthy, Christo Kirov, Matteo Grella, Amrit Nidhi, Patrick Xia, Kyle Gorman, Ekaterina Vylomova, Sabrina J. Mielke, Garrett Nicolai, Mikka Silfverberg, Timofey Arkhangel'skiy, Nataly Krizhanovsky, Andrew Krizhanovsky, Elena Klyachko, Alexey Sorokin, John Mansfield, Valts Ernštreits, **Yuval Pinter**, Cassandra L. Jacobs, Ryan Cotterell, Mans Hulden, David Yarowsky. *UniMorph 3.0: Universal Morphology*. Proceedings of the 12th Language Resources and Evaluation Conference (LREC), May 2020, pp. 3922–3931. Marseille, France.
  9. Sarthak Jain, Sarah Wiegrefe, **Yuval Pinter**, Byron C. Wallace. *Learning to Faithfully Rationalize by Construction*. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), July 2020, pp. 4459–4473. Online.
  10. **Yuval Pinter**, Cassandra L. Jacobs, Max Bittker. *NYTWIT: A Dataset of Novel Words in the New York Times*. Proceedings of the 28th International Conference on Computational Linguistics (COLING), December 2020, pp. 6509–6515. Online.
  11. **Yuval Pinter**, Cassandra L. Jacobs, Jacob Eisenstein. *Will it Unblend?* Findings of the Association for Computational Linguistics: EMNLP 2020, November 2020, pp. 1525–1535.
  12. Khuyagbaatar Batsuren et al. (88 authors). *UniMorph 4.0: Universal Morphology*. Proceedings of the 13th Language Resources and Evaluation Conference, June 2022.
  13. Ramit Sawhney, Megh Thakkar, Shrey Pandit, Ritesh Singh Soun, Sarvagya Malaviya, and **Yuval Pinter**. *CIAug: Equipping Interpolative Augmentation with Curriculum Learning*. Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, July 2022.
  14. Elazar Gershuni and **Yuval Pinter**. *Restoring Hebrew Diacritics Without a Dictionary*. Findings of the Association for Computational Linguistics: NAACL, July 2022.
  15. 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.
  16. 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.
  17. Lisa Beinborn and **Yuval Pinter**. *Analyzing Cognitive Plausibility of Subword Tokenization*. EMNLP, December 2023.
  18. **Yuval Pinter** and Michael Elhadad. *Emptying the Ocean with a Spoon: Should We Edit Models?* Findings of the Association for Computational Linguistics: EMNLP, December 2023.
  19. Carinne Cherf and **Yuval Pinter**. *BiVert: Bidirectional Vocabulary Evaluation using Relations for Machine Translation*. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), pages 2410–2420.

20. 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.
21. Anaelia Ovalle, Ninareh Mehrabi, Palash Goyal, Jwala Dhamala, Kai-Wei Chang, Richard Zemel, Aram Galstyan, **Yuval Pinter**, Rahul Gupta. *Tokenization Matters: Navigating Data-Scarce Tokenization for Gender Inclusive Language Technologies*. Findings of NAACL, June 2024.
22. Omri Uzan, Craig W. Schmidt, Chris Tanner, **Yuval Pinter**. *Greed is All You Need: An Evaluation of Tokenizer Inference Methods*. ACL, August 2024. **Outstanding Paper**.

### Refereed Articles in Scientific Journals

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.

### Refereed Articles in Workshop Proceedings

1. **Yuval Pinter**, Marc Marone, and Jacob Eisenstein. *Character Eyes: Seeing Language through Character-Level Taggers*. Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, August 2019, pp. 95–102. Florence, Italy.
2. Gal Astrach and **Yuval Pinter**. *The BGU-MeLeL System for the SIGMORPHON 2023 Shared Task on Morphological Inflection*. Proceedings of the 20th SIGMORPHON workshop on Computational Research in Phonetics, Phonology, and Morphology, July 2023.
3. Tal Kadosh, Nadav Schneider, Niranjana Hasabnis, Timothy Mattson, **Yuval Pinter**, and Gal Oren. *Advising OpenMP Parallelization via a Graph-Based Approach with Transformers*. OpenMP: Advanced Task-Based, Device and Compiler Programming. September 2023.
4. Nadav Schneider, Tal Kadosh, Niranjana Hasabnis, Timothy Mattson, **Yuval Pinter**, and Gal Oren. *MPI-RICAL: Data-Driven MPI Distributed Parallelism Assistance with Transformers*. The First Workshop on AI Assisted Software Development for HPC. November 2023.
5. Nadav Schneider, Niranjana Hasabnis, Vy A. Vo, Tal Kadosh, Neva Krien, Mihai Capotă, Abdul Wasay, Guy Tamir, Ted Willke, Nesreen Ahmed, **Yuval Pinter**, Timothy Mattson, and Gal Oren. *MPIrigen: MPI Code Generation through Domain-Specific Language Models*. AI4Sys at HPDC 2024.
6. Marco Cognetta, Tatsuya Hiraoka, Rico Sennrich, Naoaki Okazaki and **Yuval Pinter**. *An Analysis of BPE Vocabulary Trimming in Neural Machine Translation*. The 5th Workshop on Insights from Negative Results in NLP, June 2024.
7. Reem Harel, Yair Elboher, and **Yuval Pinter**. *Protecting Privacy in Classifiers by Token Manipulation*. Fifth Workshop on Privacy in Natural Language Processing at ACL, August 2024.

### Published Abstracts, Scientific Reports and Technical Papers

1. Eugene Agichtein, David Carmel, Donna Harman, Dan Pelleg, and **Yuval Pinter**. *Overview of the TREC 2015 Live QA Track*. TREC, November 2015.
2. Eugene Agichtein, David Carmel, Donna Harman, Dan Pelleg, and **Yuval Pinter**. *Overview of the TREC 2016 Live QA Track*. TREC, November 2016.
3. Asma Ben Abacha, Eugene Agichtein, **Yuval Pinter**, and Dina Demner-Fushman. *Overview of the Medical Question Answering Task at TREC 2017 LiveQA*. TREC, November 2017.
4. Afra Alishahi, Yonatan Belinkov, Grzegorz Chrupala, Dieuwke Hupkes, **Yuval Pinter**, Hassan Sajjad. *Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*. Association for Computational Linguistics, November 2020.
5. Jasmijn Bastings, Yonatan Belinkov, Emmanuel Dupoux, Mario Giulianelli, Dieuwke Hupkes, **Yuval Pinter**, Hassan Sajjad. *Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*. Association for Computational Linguistics, November 2021.
6. Re'em Harel, **Yuval Pinter**, Gal Oren. *Learning to parallelize in a shared-memory environment with transformers*. Proceedings of the 28th ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming, February 2023.

### Patents

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.

## GRANTS

### Competitive Grants

- 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 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**

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

**Research Students, graduated**

1. Carinne Cherf, MSc. (2021–2024)
2. Peter Kogan, MSc. (2021–2024)
3. Shaked Yehezkel, MSc. (joint with Omer Levy, TAU) (2021–2022)

**Research Students, current**

1. Shira Darchi, Ph.D (joint with Talya Sadeh) (s. 2023)
2. Re'em Harel, Ph.D (s. 2022)
3. Adi Cohen, MSc. (s. 2024)
4. Rotem Brillant, MSc. (s. 2024)
5. Bar Block, MSc. (s. 2024)
6. Or Kadosh, MSc. (s. 2024)
7. Bar Gazit, MSc. (s. 2023)
8. Omri Uzan, MSc. (s. 2023)
9. Yair Elboher, MSc. (s. 2022)
10. Nadav Schneider, MSc. (joint with Gal Oren, Technion and Dan Vilenchik, BGU EE) (s. 2022)
11. Tal Kadosh, MSc. (joint with Gal Oren) (s. 2022)
12. Tamar Levy, MSc. (joint with Jonathan Berant, TAU) (s. 2022)
13. Gal Astrach, MSc. (s. 2022)
14. Ophir Harouche, MSc. (joint with Yoav Goldberg, BIU) (s. 2021)

**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).