

The Landscape of Computational Linguistics

Jordan Boyd-Graber

University of Maryland

November 3, 2024

Conferences

ML/AI



- ICML
- NeurIPS
- AAAI
- IJCAI
- AISTATS
- ICLR

Data Mining



- WWW
- KDD
- ICWSM

Language



- ACL
- NAACL
- EMNLP
- Coling
- CoNLL
- LREC

New Kids on the Block

- ICLR
- COLM
- *Sem

Journals

- Machine Learning Journal (MLJ)
- Computational Linguistics (CL)
- Journal of Machine Learning Research (JMLR)
- Transactions of Computational Linguistics (TACL)



JMLR

Journals

- Machine Learning Journal (MLJ)
- Computational Linguistics (CL)
- Journal of Machine Learning Research (JMLR)
- Transactions of Computational Linguistics (TACL)
- Findings of the Association for Computational Linguistics (Findings)



JMLR

Preprints



Cornell University

arXiv.org

- ArXiv is best known, but also OpenReview, etc.
- Gets your research out in the world (some journalists look for preprints)
- Not allowed to do it for some conferences!
- **Not a peer-reviewed publication!** Healthy skepticism is warranted for things you read on preprint servers.

Sub-areas

Computational Social
Science
Dialogue
Discourse
Ethics
Information Extraction
Information Retrieval
Interpretability

Language Grounding
Linguistic theories
Machine Learning
Machine Translation
Applications
Phonology
Question Answering

Resources
Semantics
Sentiment
Speech
Summarization
Syntax

Affinity Groups

- Queer in AI
- Black in AI
- Women in Machine Learning

The Landscape of Computational Linguistics

Jordan Boyd-Graber

University of Maryland

Jobs

Sub-areas

Academia

Industry

Gov Labs

Sub-areas

Academia

Lecturer

Research Professor

Professor

Industry

Gov Labs

Sub-areas

Academia

Lecturer
Research Professor
Professor

Industry

Software Engineer
Research Scientist

Gov Labs

Sub-areas

Academia

Lecturer
Research Professor
Professor

Industry

Software Engineer
Research Scientist

Gov Labs

Researcher
Program Manager

Where Does Funding come From?

- In Industry
 - ▶ Convince your boss!
- Academia
 - ▶ Corporate gifts
 - ▶ “Blue sky” research (NSF/NIH)
 - ▶ Deliverable-based research (DARPA/IARPA/Industry)
- In Government Labs
 - ▶ Customers / internal grants

Pros and Cons of Industry vs. Academia

- Industry

PRO

Takehome salary

Perks

Data

More specialization

CON

Interns

Need to justify research

Prepublication Review

- Academia

PRO

Flexibility

Students

Freedom

CON

Fight for funding

Paperwork

Pros and Cons of Industry vs. Academia

- Industry

PRO

Takehome salary

Perks

Data

More specialization

CON

Interns

Need to justify research

Prepublication Review

- Academia

PRO

Flexibility

Students

Freedom

CON

Fight for funding

Paperwork

Pros and Cons of Industry vs. Academia

- Industry

PRO

Takehome salary

Perks

Data

More specialization

CON

Interns

Need to justify research

Prepublication Review

- Academia

PRO

Flexibility

Students

Freedom

CON

Fight for funding

Paperwork

Pros and Cons of Industry vs. Academia

- Industry

PRO

Takehome salary

Perks

Data

More specialization

CON

Interns

Need to justify research

Prepublication Review

- Academia

PRO

Flexibility

Students

Freedom

CON

Fight for funding

Paperwork

Pros and Cons of Industry vs. Academia

- Industry

PRO

Takehome salary

Perks

Data

More specialization

CON

Interns

Need to justify research

Prepublication Review

- Academia

PRO

Flexibility

Students

Freedom

CON

Fight for funding

Paperwork

Pros and Cons of Industry vs. Academia

- Industry

PRO

Takehome salary

Perks

Data

More specialization

CON

Interns

Need to justify research

Prepublication Review

- Academia

PRO

Flexibility

Students

Freedom

CON

Fight for funding

Paperwork

Why not a startup?

- Startups often don't navigate immigration as well as established firms
- More control of the vision
- Unless goal is accuhire, won't be getting much research done
- Need to solve a problem, not just have cool tech

Jobs Outside of Research

- Data scientist
- Data journalist
- Consultant
- Data wrangler
- Policy advisor

The Landscape of Computational Linguistics

Jordan Boyd-Graber

University of Maryland

Paths

Kinds of contributions in computational linguistics

1. The Theorist
2. The Problem Definer
3. The Problem Solver
4. The Boundary Redefiner
5. The Resource Creator
6. The Leader
7. The Facilitator

Kinds of contributions in computational linguistics

1. The Theorist

Fillmore, 1991

Armchair linguistics does not have a good name in some linguistics circles. A caricature of the armchair linguist is something like this. He sits in a deep soft comfortable armchair, with his eyes closed and his hands clasped behind his head. Once in a while he opens his eyes, sits up abruptly shouting, “Wow, what a neat fact!,” grabs his pencil, and writes something down. Then he paces around for a few hours in the excitement of having come still closer to knowing what language is really like.

2. The Problem Definer
3. The Problem Solver
4. The Boundary Redefiner
5. The Resource Creator
6. The Leader
7. The Facilitator

Kinds of contributions in computational linguistics

1. The Theorist
2. The Problem Definer
 - ▶ Sentiment Analysis: Turney (2001), Pang and Lee (2002)
 - ▶ Probabilistic Topic Modeling: Thomas Hoffman (1999)
 - ▶ Word Embeddings: Mikolov et al. (2013)
 - ▶ **Be First**
3. The Problem Solver
4. The Boundary Redefiner
5. The Resource Creator
6. The Leader
7. The Facilitator

Kinds of contributions in computational linguistics

1. The Theorist
2. The Problem Definer
3. The Problem Solver
 - ▶ Always win the leaderboard
 - ▶ Funding organized around this
 - ▶ **Be Best**
4. The Boundary Redefiner
5. The Resource Creator
6. The Leader
7. The Facilitator

Kinds of contributions in computational linguistics

1. The Theorist
2. The Problem Definer
3. The Problem Solver
4. The Boundary Redefiner
 - ▶ Using Gibbs sampling (from physics)
 - ▶ Using good statistical tests
 - ▶ Paying attention to ethics
5. The Resource Creator
6. The Leader
7. The Facilitator

Kinds of contributions in computational linguistics

1. The Theorist
2. The Problem Definer
3. The Problem Solver
4. The Boundary Redefiner
5. The Resource Creator
 - ▶ Unique to ML/NLP!
 - ▶ Many people do this at small scale
 - ▶ Really hard to do this at scale so the contribution lasts
 - ▶ Feifei Li (Stanford)
 - ▶ Martha Palmer (Colorado)
 - ▶ Linguistic Data Consortium (U Penn)
6. The Leader
7. The Facilitator

Kinds of contributions in computational linguistics

1. The Theorist
2. The Problem Definer
3. The Problem Solver
4. The Boundary Redefiner
5. The Resource Creator
6. The Leader
 - ▶ Build a team
 - ▶ Get funding
 - ▶ Make them play well together
7. The Facilitator

Kinds of contributions in computational linguistics

1. The Theorist
2. The Problem Definer
3. The Problem Solver
4. The Boundary Redefiner
5. The Resource Creator
6. The Leader
7. The Facilitator
 - ▶ Toolkits
 - ▶ Running a conference
 - ▶ Mentoring others

Schools



- Stanford
- UW
- Columbia
- CMU
- MIT
- TTI/Chicago

Schools



- Maryland
- Stanford
- UW
- Columbia
- CMU
- MIT
- TTI/Chicago

Companies: 1990s

Twentieth Century

- 1990s
 - ▶ Microsoft
 - ▶ AT&T
- 2000s
 - ▶ Google
 - ▶ Microsoft
 - ▶ Yahoo!

Twenty-First Century

- 2010s
 - ▶ Google
 - ▶ Facebook
 - ▶ Amazon
 - ▶ Microsoft

What about now?

Too early to say!

What about now?

Too early to say!

- Huggingface
- NVIDIA
- Cohere
- Anthropic
- OpenAI

Sects

- Max-Margin
- Theoretical
- Deep
- Bayesian
- Reinforcement