



Brennan Dell - Flash Talk

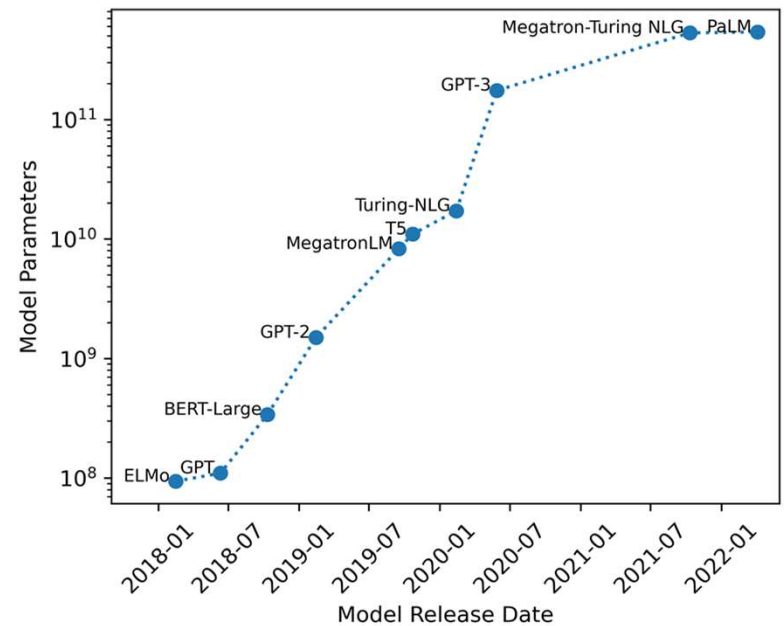
Background

- Assistant Professor of Linguistics at UVM
- BS in Computer Science from the Rochester Institute of Technology
- PhD in Linguistics from the University of Illinois Urbana-Champaign
- Research interests
 - Machine translation
 - Translation evaluation
 - Interpretable natural language processing
 - NLP in context



Evaluation in Natural Language Processing

- NLP systems have increased in size and general abilities
- However, underlying abilities have become less clear
- External evaluation is crucial
- Fine-grained evaluation > one-dimensional evaluation





Challenges of Translation Evaluation

- How can we objectively compare the performance of different MT systems?
- Challenges:
 - Periphrasis
 - Translations are not one-to-one
 - Multidimensionality
 - Is the meaning correct?
 - Is the target sentence syntactically well-formed?
 - Is the target sentence appropriate for the document type and social setting?
 - Etc...
 - Subjectivity
 - Humans can disagree on translation quality



My recent work: Translation of Evidentiality

- Evidentiality: grammatical indication of a speaker's knowledge
 - Direct: certainty, witnessed
 - Indirect: hearsay, inference, doubt
- Findings:
 - Translation models seem to ignore Turkish evidentiality into English
 - Translation models seem to avoid evidentiality translated into Bulgarian
- Further questions:
 - Are other underrepresented phenomena similar?
 - Do humans perceive this difference in translation?

ağaç düş -müş

tree fall -EVID

“(Apparently) the tree fell”

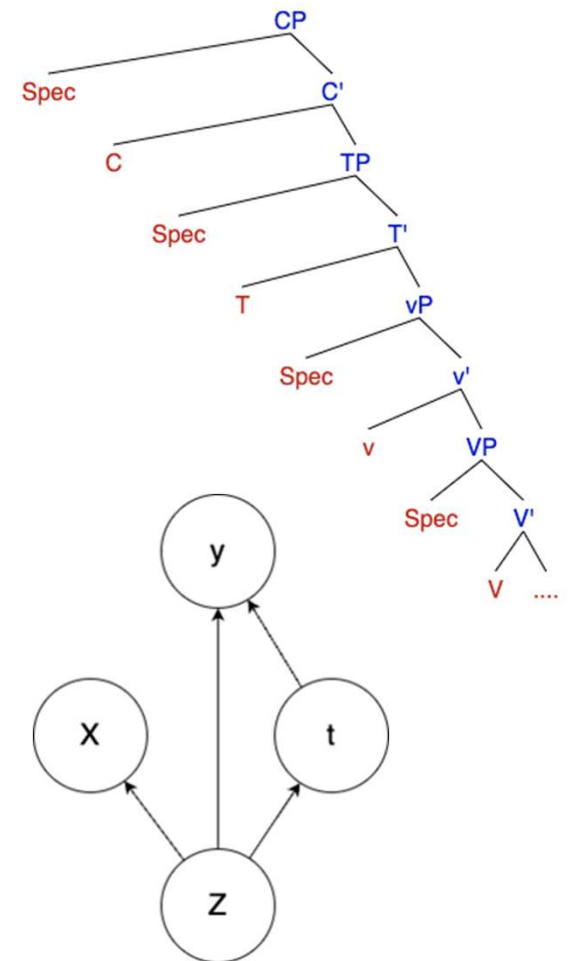
ağaç düş -tü

tree fall -PST.3P

“The tree fell”

Future work: Evaluation in context

- How do we measure the abilities of NLP systems?
 - Where do we even start?
- Use findings from other disciplines
 - Linguistic theory, psycholinguistics - study language structure and process
 - Medicine, economics - study causality in complex systems
- Start with the context and the stakeholders
 - Variety of world's languages and linguistic features
 - Evaluate for the cost



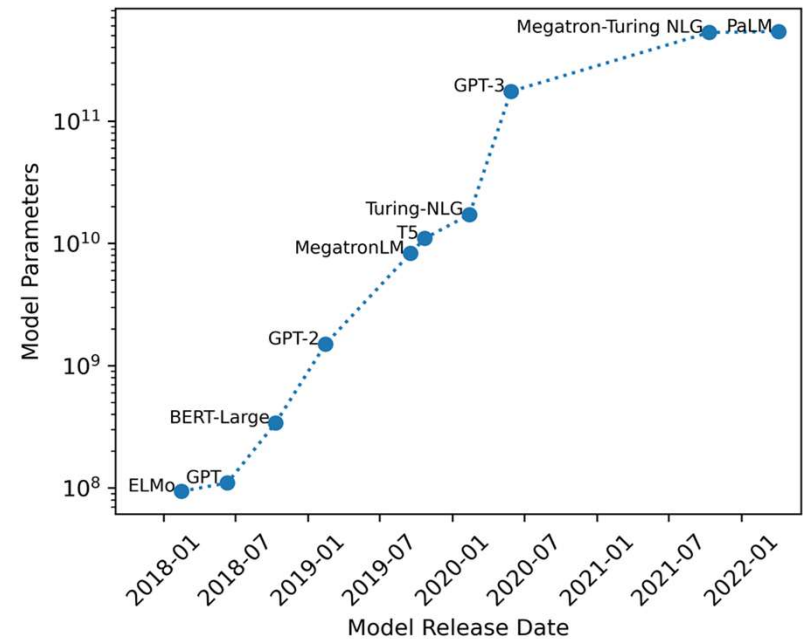


Grammatical Evaluation for NLP

- Focus on “challenge sets”
 - Data sets which target a specific challenging feature
 - Identify “pinch point” phenomena
- Major area of recent interest in machine translation ([Kocmi et al. 2024](#))
- Use language acquisition and psycholinguistic research

Evaluation for Efficiency

- Cost is rarely a component of evaluation
 - Cost is financial, computational, environmental
- Language models have grown exponentially in size
- Cost limits developments to large companies
 - Limits equity in research
- Future research: include cost and efficiency as a component of evaluation



Data source:

<https://observablehq.com/@sorami/sizes-of-large-language-models>



Social Inclusion in NLP

- Evaluate NLP systems' role in socio-technical system
- Translation is not just decipherment
 - Communication across different cultures
- Goals
 - Reduce social harms
 - Meet the needs of the users
 - Gain insights from the users



Contact Info

Email: brennan.dell@uvm.edu

Office: 522 Waterman