

Internship at George Mason University

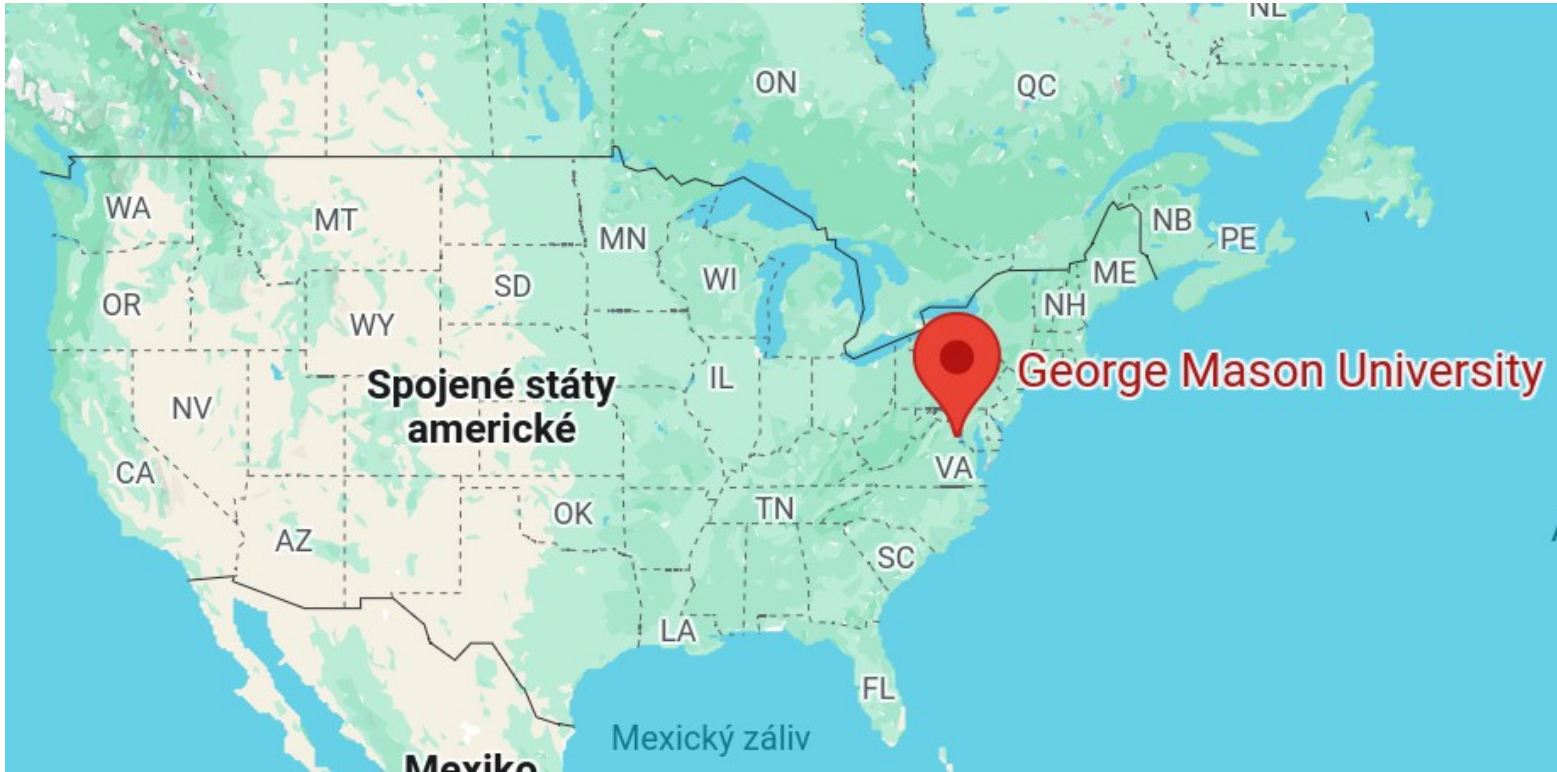
Hana Hledíková
ÚFAL outdoor seminar 2024



Structure of the presentation

- Internship at a university – environment, people, research
- Steps and timeline
- Research projects
- Overall experience – pros and cons
- Travel photos :)

George Mason University



George Mason University



George Mason University



George Mason University



James P. Bleviss

head of Linguistics Department



Geraldine Walther

Assistant Professor
head of the *Computational
Linguistics Lab*

Research

- Main goal: **consult dissertation research** with prof. Blevins
 - complexity of the verbal lexicon
 - comparative + data-based perspective
 - distribution of morphological families and morphemic complexity across frequency spectrum
 - interface of word-formation and syntax

Research

- prof. Blevins:
 - primary focus: structure and complexity of grammatical systems
 - **computational methods** to address **theoretical questions**
- language as a complex system
- information-theoretic approach
- Word and Paradigm morphology
- data sparsity and the consequences for paradigmatic structure

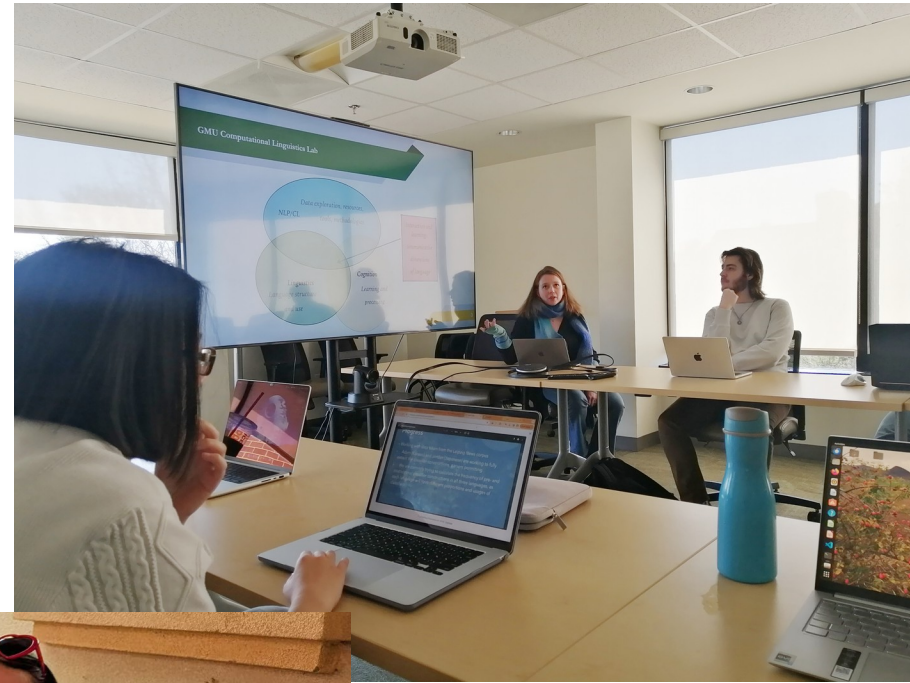
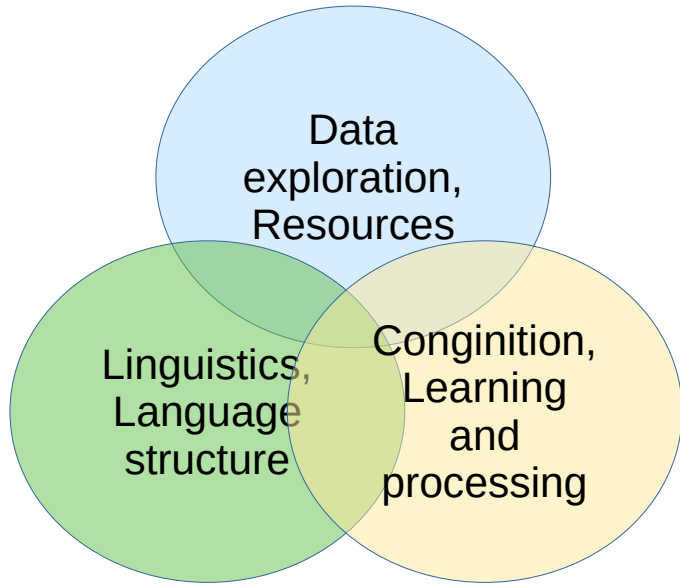




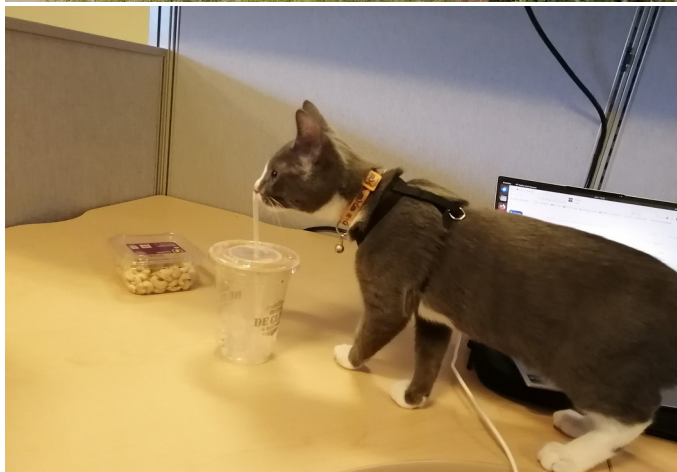
Steps + timeline

- 31 March 2023: first e-mail
- 31 March 2023: first (positive) response
- March–May 2023: agreement on date and details
- 25 May 2023: started visa process
- 13 October 2023: got visa
- 8 January 2024: arrived in US
- 7 July 2024: returned to Prague

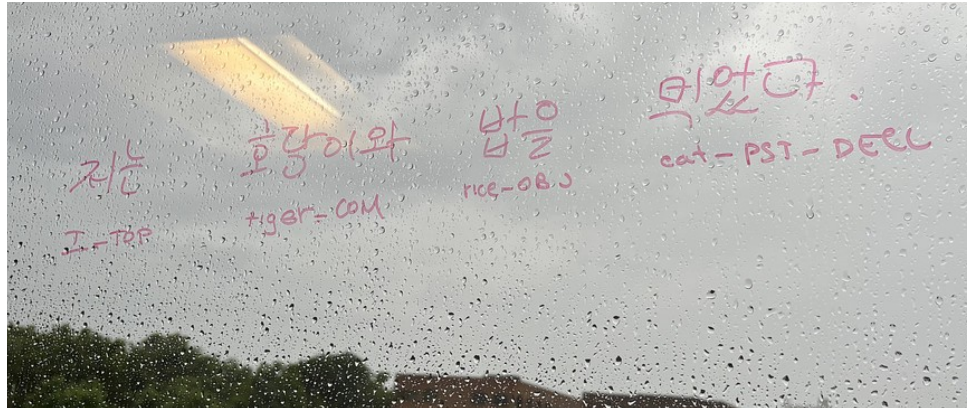
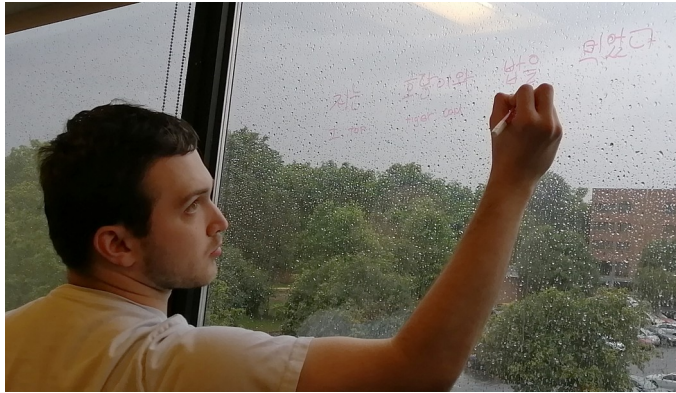
Computational Linguistics Lab



Computational Linguistics Lab



Research: Korean verbs



MA thesis:

In Context Verb Disambiguation: A Korean Case Study with Reference to Czech

Processing perspective: prediction in context, entropy smoothing

- Nouns: gender, pre-nominal classifiers, adjectives
- Verbs: formal characteristics of arguments
- **Korean**: verb-final → pre-verbal dependents ~ pre-nominal classifiers/adjectives
- **Czech**: flexible word order – comparison

Research: Kulung Documentation



Kulung: endangered language in Nepal

Goals:

- **Collect data**
- **Develop tools** – POS tagger, automatic transcription tool, machine translation into/from English
- **Answer linguistic questions**

Tentative plan:

- ? ComputEL-8 Workshop on the Use of Computational Methods in the Study of Endangered Languages
- ? First Workshop on Language Models for Low-Resource Languages at COLING

Research: my dissertation topic

1) Effects on verb processing in Czech

- **frequency**
- **paradigmatic dimension:**
 - morphological family size
- **syntagmatic dimension:**
 - number and forms of (preceding) arguments

→ submission to COLING 2025

2) Plan for future joint research:

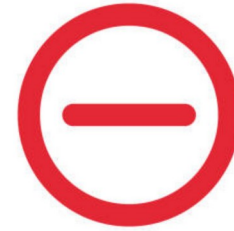
- typological, corpus-based
- inference of **inflection classes + w-f paradigms** from **sparse data**

→ submission to Complexity in Language Sciences Meeting in Paris

Summary of overall experience



- openness, easy integration into lab
- community aspect, useful meetings
- AC in the office
- interesting classes
- location
- travelling opportunities



- time scheduling
- no car
- expensive area
- AC in the office

Washington DC



2024. 03. 22. 10:55

Virginia



West Virginia → South Carolina



Florida



2024. 03. 04. 18:35



2024. 03. 01. 10:24