



ÚSTAV  
BOHEMISTICKÝCH STUDIÍ  
Filozofická fakulta  
Univerzita Karlova



FILOZOFICKÁ FAKULTA  
Univerzita Karlova

Ústav bohemistických studií

Jsou žáci na podprahové  
úrovni schopni produkovat  
různorodé texty?

Multidimenzionální přístup k  
češtině jako cizímu jazyku

Adrian Jan Zasina

Ostrava, 25. 1. 2023

# Poděkování

---

Tato práce vznikla za podpory projektu Mezinárodní mobility výzkumných, technických a administrativních pracovníků na UK č. CZ.02.2.69/0.0/0.0/18\_053/0016976.



EVROPSKÁ UNIE  
Evropské strukturální a investiční fondy  
Operační program Výzkum, vývoj a vzdělávání



# Úvodem

---

- Současné empirické studie provedené na angličtině nerodilých mluvčích potvrzují, že pokročilí žáci jsou schopni produkovat různorodé texty (Biber et al., 2016; Qiu et al., 2022; Staples et al., 2018)
- Zkoumání psané produkce žáků na podprahové úrovni není běžné (Kim & Nam, 2019)
- Použití multidimenzionálního modelu češtiny ke zkoumání podobnosti a rozdílu v rámci nerodilých českých textů

# Výzkumné otázky

---

- Vytvářejí žáci na podprahové úrovni češtiny různorodé texty v rámci různých zadání s ohledem na jazykovou variabilitu?
  1. Zda jednotlivá zadání projevují rozdílné charakteristiky v psané produkci žáků na podprahové úrovni?
  2. Lze identifikovat nějaké podobnosti mezi rodilými a nerodilými texty z hlediska registrů v češtině?

# Multidimenzionální analýza

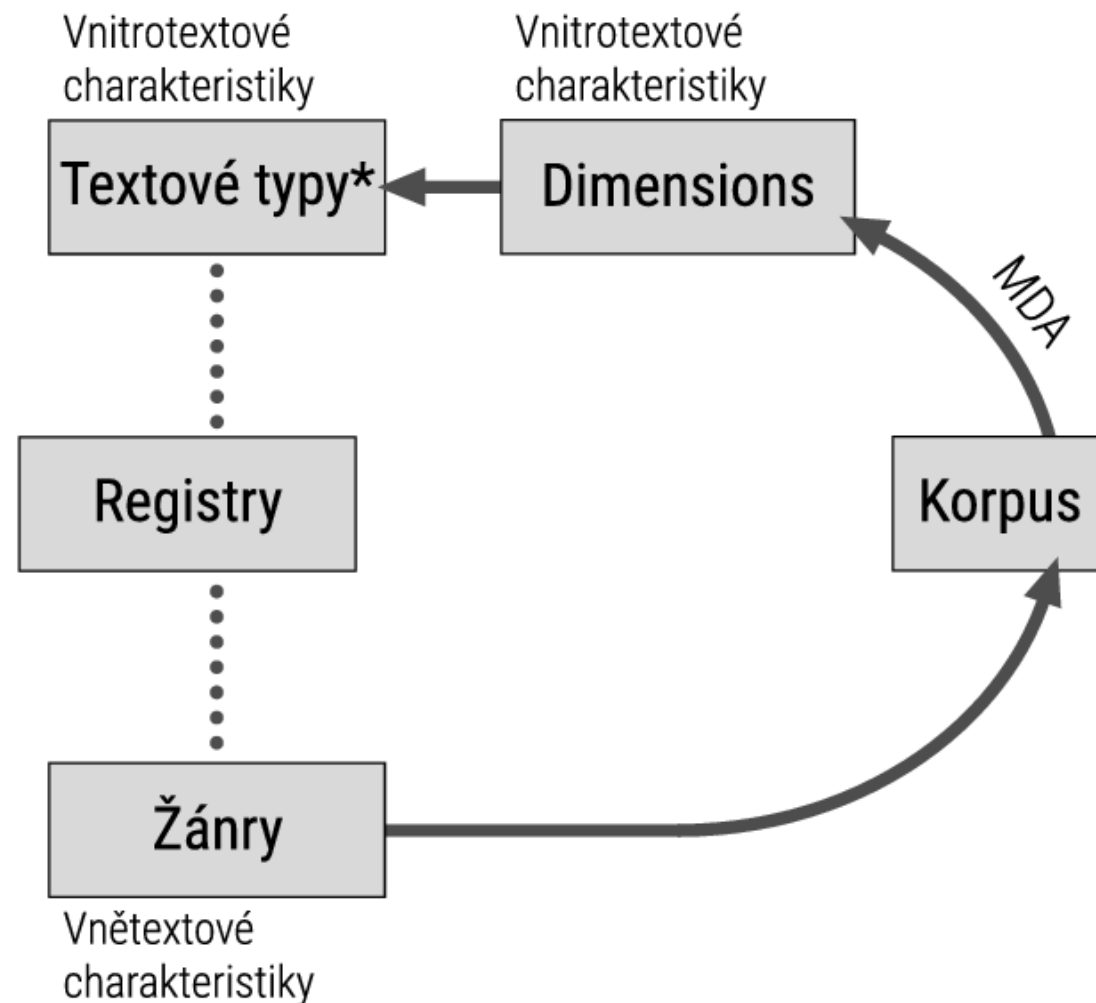
# Multidimenzionální analýza(MDA)

---

- MDA se snaží uchopit rozsah variace v jazyce na základě souvýskytu jazykových rysů
- Jedná se o funkční variability
- Poprvé provedená pro rodilou angličtinu D. Biberem (1988, 1989)
- Později aplikovaná na mnoho dalších jazyků: španělštinu, brazilskou portugalsčinu, korejštinu, taiwanštinu, tuvalštinu, (cf. Biber, 2014)

# Fáze MDA

1. Sestavení korpusu
2. Jazykové rysy: operacionalizace & extrakce
3. Statistické vyhodnocení (faktorová analýza)
4. Interpretace výsledků



# Žánr vs. registr



Vánoční cukroví

Zdroj: <https://pxhere.com/en/photo/672515>

# České dimenze variability

---

1. dynamický (+) vs. statický (-)
2. spontánní (+) vs. připravený (-)
3. vyšší (+) vs. nižší (-) stupeň koheze
4. polytematický (+) vs. monotematický (-)
5. vyšší (+) vs. nižší (-) míra explicitní adresnosti
6. obecný (+) vs. konkrétní (-)
7. prospektivní (+) vs. retrospektivní (-)
8. postojovost (+) vs. faktualnost (-)

# České dimenze variability

---

1. dynamický (+) vs. statický (-)
2. spontánní (+) vs. připravený (-)
3. vyšší (+) vs. nižší (-) stupeň koheze
4. polytematický (+) vs. monotematický (-)
5. vyšší (+) vs. nižší (-) míra explicitní adresnosti
6. obecný (+) vs. konkrétní (-)
7. prospektivní (+) vs. retrospektivní (-)
8. postojovost (+) vs. faktualnost (-)

# Data a metodologie

# Data

---

## Žákovské texty

- Studenti **prvního ročníku Bohemistiky pro cizince** na Jagellonské univerzitě v Krakově
- Sběr dat provedený během speciálního kurzu *Gramaticko-lexikální cvičení pro bohemisty*
- **16** studentů (14 žen, 2 muži)
- Úroveň: **A2**
- Věk: **19-24**

# Zadání

---

1. Neformální dopis {DOP}
2. Popis místa {MIS}
3. Reflexe (argumentace) na téma výhod a nevýhod současných médií {MED}
4. Vyprávění o světě v roce 3000 {300}

# Zadání

---

1. {DOP} – neformální dopis může vyjadřovat osobní postoj (*Jaká česká jídla jste ochutnal/-a?*), komentovat události (*návštěva Prahy*), poskytovat informace o faktech (*Jak dlouho tady budete?*) a může používat více kauzálních frází k popisu činností (*Co v Praze děláte?*)
2. {MIS} – popis divného místa má tendenci k použití minulého času, popisu události (*Kde a kdy to bylo?*) a situace (*Co bylo na tom divného?*)

# Zadání

---

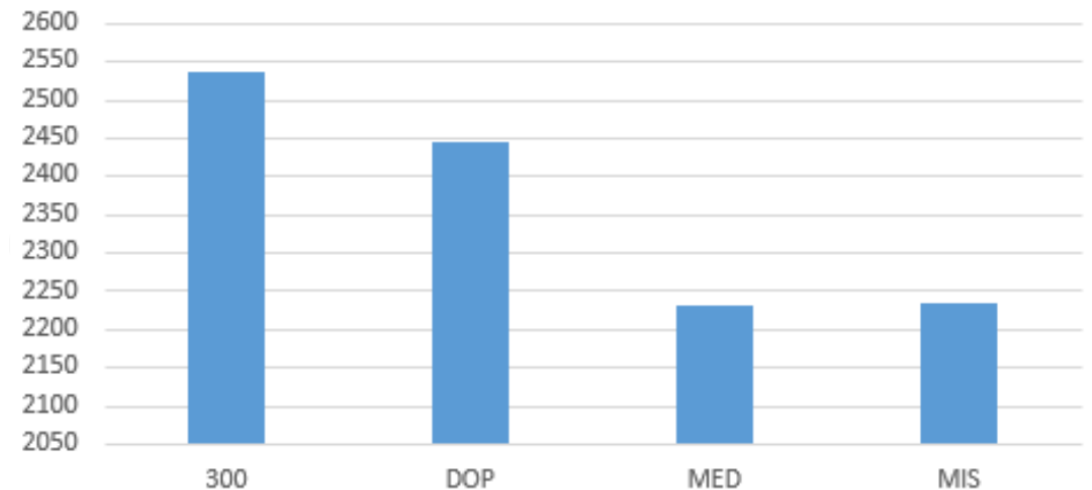
3. {MED} – argumentace je převážně úvahou o kladech a záporech dnešních médií (*Vyjádřete svůj kladný nebo záporný vztah k uvedené otázce*); může poskytovat vlastní nápady a argumenty na podporu vlastního názoru (*Mohou být současná média ohrožením pro lidskost?*)
4. {300} – vyprávění je o faktech v budoucnosti (*Jaká bude politická situace na světě?*); pisatel popisuje vlastní představivost (*Jak se změní příroda?*) a abstrakci (*Bude možný let na Mars?*)

# Žákovské texty

---

- Napsané mimo hodinu v rámci DÚ
- 62 texty
- 150 slov na text
- 125-203 rozsah slov
- Texty byly zpracovány v nástroji TEITOK (Janssen, 2016)
- Velikost korpusu: 9,447 slov

Graf 1. Počet slov pro každé zadání v korpusu

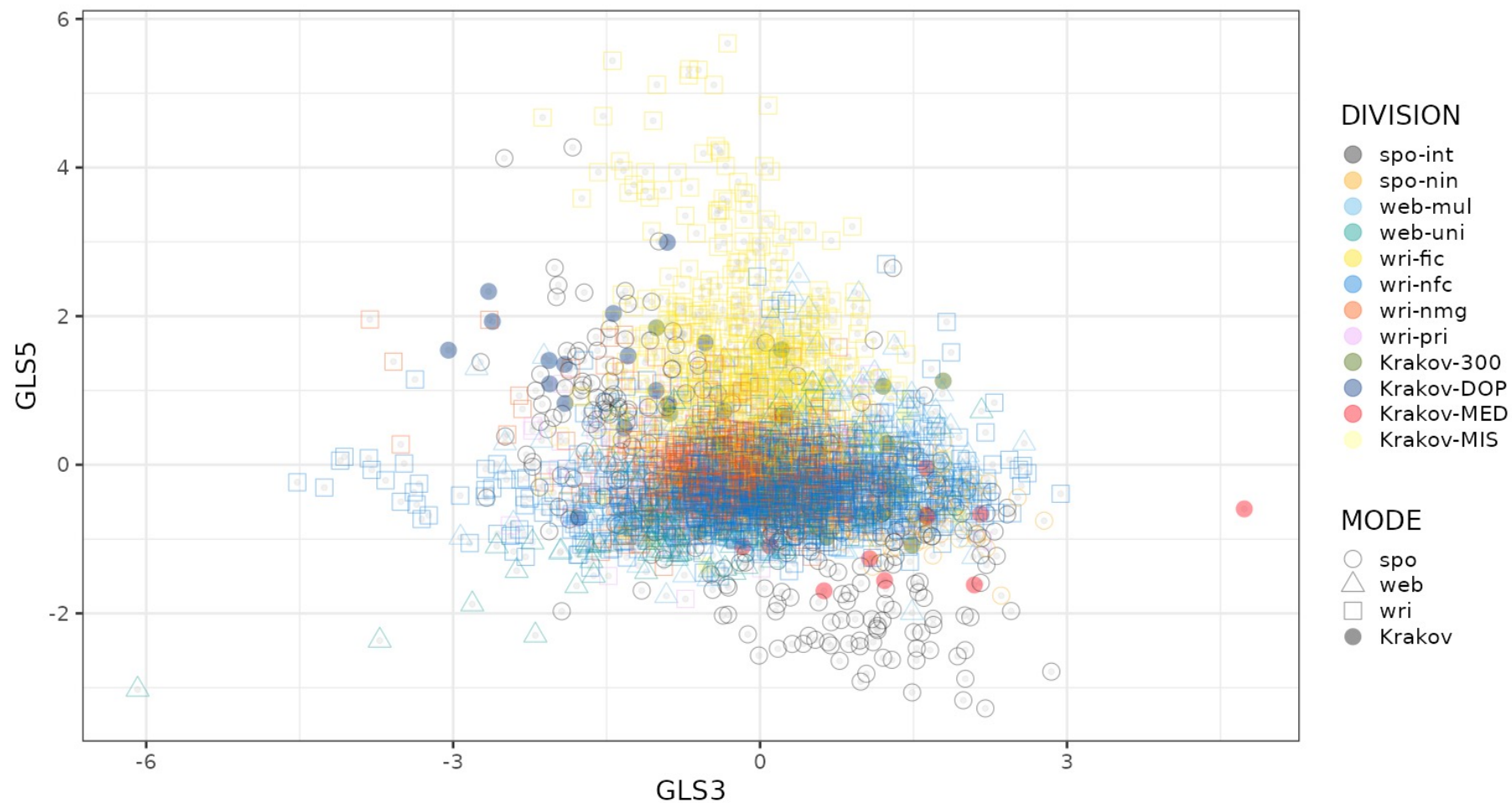


# Operacionalizace

---

1. Anotace dat stejným způsobem jako původní česká MDA
2. Použití 122 jazykových rysů
3. Faktorová analýza (Biber, 1988, p. 79)
4. Projekce žákovských textů na 8-dimenzionální prostor
5. Interpretace žákovských textů v rámci 8-dimenzionálního prostoru

# 2-dimenzionální prostor



# Charakteristika zadání

# Srovnání mediánů faktorových skóre

| DIM1 |      | DIM2  |       | DIM3  |       | DIM4 |      |
|------|------|-------|-------|-------|-------|------|------|
| DOP  | MIS  | DOP   | MIS   | DOP   | MIS   | DOP  | MIS  |
| 0,84 | 0,95 | 0,20  | 0,38  | -1,61 | -0,52 | 0,81 | 0,85 |
| 0,77 | 0,67 | -0,05 | -0,54 | 1,22  | 0,39  | 0,30 | 0,30 |
| MED  | 300  | MED   | 300   | MED   | 300   | MED  | 300  |

| DIM5  |       | DIM6  |       | DIM7 |      | DIM8  |       |
|-------|-------|-------|-------|------|------|-------|-------|
| DOP   | MIS   | DOP   | MIS   | DOP  | MIS  | DOP   | MIS   |
| 1,38  | -0,56 | -0,51 | 0,55  | 0,86 | 1,11 | -0,69 | -0,40 |
| -0,68 | 0,23  | 0,68  | -1,65 | 2,82 | 3,47 | 0,07  | -0,75 |
| MED   | 300   | MED   | 300   | MED  | 300  | MED   | 300   |

# Neformální dopis {DOP}

---

- Nejnižší stupeň koheze (DIM3) :
  - číslovky (základní: *dva, tři dny*, řadové: první)
  - nízký výskyt: *korelativa*, *jmenný přísudek substantivní*, *vztažné věty typu který*, *inventář zájmen*, *inventář spojek*, *kondicionál*, *dativ substantiv*, *komplementace slovesa*, a frekventované *n-gramy*
  - naprostý nedostatek: *morfologická konkurence koncovek -é vs. -i* a *infinitiv na -ci*
- Nejvyšší míra explicitní adresnosti (DIM5):
  - všechny typy *otázek* (*Byla jsi někdy v Praze?*), *slovesa ve 2. osobě* (*jsi, máš*), *otázky doplňovací* (*Jak se máš?*) a *zájmena ve 2. osobě* (*ti, tvůj*)
- Bohatá slovní zásoba (DIM4):
  - *toponyma* (*Praha, Česká republika, Evropa*), *nominativ* a *akuzativ* (*Včera jsem navštívila Pražský hrad*)
- Koncentrace informací (DIM6)

# Popis místa {MIS}

---

- Nejdynamičtější zadání (DIM1) :

- časté užití **minulého času** (*To bylo divné místo*)
- **slovesa**, **verba finita**, **indikativ** a **slovesný vid – perfektum**

- Nejspontánnější (DIM2):

- mírně vyšší skóre pro **ukazovací zájmena** (*Když o tom mluvím...*), **adverbia vyjadřující určení místa či směru** (*byly tam odpadky*) a **polyfunkční to** (*to je moje univerzita*)

- Nejpolytematičtější (DIM4):

- **toponyma** (*v Krakově, v Polsku*)

- Mírně nižší stupeň koheze (DIM3)

- Nižší míra explicitní adresnosti (DIM5)

# Argumentace {MED}

---

- Nejvyšší stupeň koheze (DIM3) :
  - **přítomný čas** (*Dnes media jsou hlavním zdrojem informací*), **dativ** (*Internet pomáhá při učení nejen žákům nebo studentům, ale také dospělým lidem*) a **komplementace slovesa** (*Myslím že média namají jen špatných vlastností*)
- Nejnižší míra explicitní adresnosti (DIM5):
  - nedostatek: **otázky doplňovací**, **zájmena ve 2. osobě**, **vykřičník**, **vokativ**, **imperativ**
- Středně prospektivní (DIM7):
  - **přítomný čas** a užití **modální substantiva**, **adjektiva**, **adverbia** a **partikule** (*chceme, jistá, samozřejmě, třeba*)
- Obecné vyjadřování (DIM6):
  - **koordinace** (*Mohou učit a rozšířit horizonty*), **sémanticky vyprázdněná adjektiva** (*celý, jednoduchý, volný*) a **konjunkce** (*Ale myslím, že nejsou vždy důvěryhodné*)

# Vyprávění {300}

---

## ■ Nejpřípravenější (DIM2):

- **nominative-akuzativ** (*svět, přírodu*), **substantiva obecně** (*budoucnost, důvěra, cestování*)
- nízký výskyt: **kontaktové výrazy**, **expresivní částice vyjadřující hodnocení, postoj, modalitu**, úžení -é > -í/ý v koncovce, **polyfunkční to**, **sémanticky vyprázdněná adverbia**
- naprostý nedostatek : **interjekce**, **přehláska ý > ej v kořeni** a **v koncovce**, **opakování identických slov**

## ■ Nejkonkrétnější vyjadřování (DIM6):

- **číslovky** (*několik, 3000, druhou*) a **klastry substantiv** (*Jižní Americe, úroveň moří*)

## ■ Nejprospektivnější (DIM7):

- **slovesa ve futuru** (*Nebude tam léto nebo jaro, Nebude tam žádná důvěra*) a **jmenný přísudek adjektivní** (*je to možné, to bude nepotřebné*)

## ■ Nejvíce faktuální (DIM8)

# Registrová vzdálenost

# Registry v češtině

---

## ■ Statické registry:

1. **Analýza**: statický monotematický
2. **Popularizace**: statický polytematický obecný
3. **Žurnalistika**: statický nevyhraněný
4. **Fakta**: statický polytematický konkrétní
5. **Argumentace**: statický kohezní

## ■ Dynamické registry:

1. **Anketa**: dynamický neadresný
2. **Konverzace**: dynamický spontánní
3. **Komentář**: dynamický postojový
4. **Scénář**: dynamický adresný
5. **Narace**: dynamický retrospektivní

# Registrová vzdálenost žákovských textů

---

| REGISTER                 | DOP  | MIS  | MED  | 300  |
|--------------------------|------|------|------|------|
| 1. ANALYSIS              | 4.88 | 4.24 | 5.35 | 5.53 |
| 2. POPULARISATION        | 3.45 | 3.34 | 4.19 | 4.93 |
| 3. QUESTION<br>ANSWERING | 5.18 | 4.09 | 3.68 | 4.98 |
| 4. CONVERSATION          | 4.22 | 4.32 | 5.81 | 5.96 |
| 5. COMMENTARY            | 3.38 | 2.90 | 3.35 | 3.96 |
| 6. JOURNALISM            | 3.53 | 3.20 | 4.21 | 5.01 |
| 7. SCREENPLAY            | 3.55 | 5.02 | 6.20 | 5.16 |
| 8. FACTS                 | 3.66 | 4.00 | 5.30 | 4.65 |
| 9. NARRATION             | 3.59 | 3.62 | 4.95 | 5.71 |
| 10. ARGUMENTATION        | 4.61 | 3.71 | 3.97 | 5.21 |

# Registrová vzdálenost - shrnutí

---

- Všechna zadání silně korelují s komentářem, který je dynamický a extrémně postojový
- Nejvágnějším zadáním je {DOP} a následně {MIS}
- Žákovská argumentace {MED} je vysoce dynamická kvůli nízké jazykové úrovni studentů
- {300} má nejblíž ke komentáři, který je registrem sdružujícím především webové texty a soukromou korespondenci; tohoto typu texty jsou vysoce dynamické a prospektivní

# Závěry a výhled do budoucna

# Interpretace

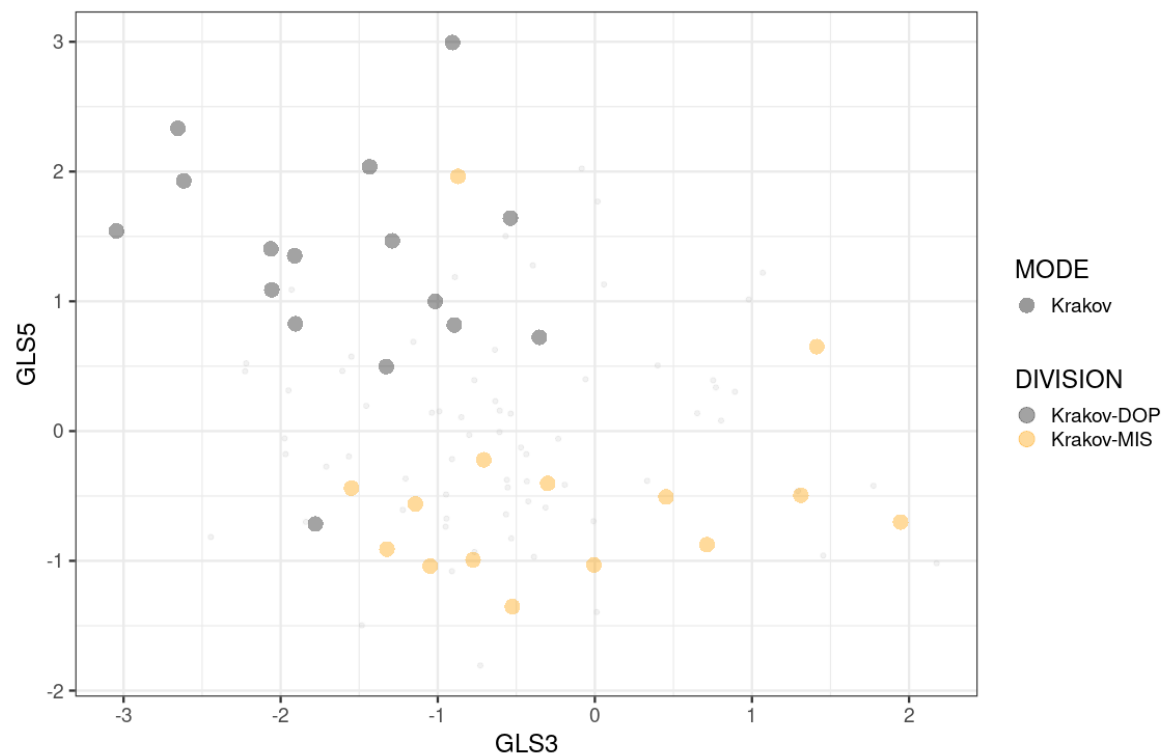
---

Čtyři zvolené úkoly jsou z hlediska funkce různorodé.

- Dopis má nejnižší stupeň koheze a nejvyšší míru explicitní adresnosti
- Popis místa je spontánní a polytematický
- Argumentační esej má nejvyšší míru koheze a nejnižší míru explicitní adresnosti
- Vyprávění nejvíce konkretizuje informace a je nejprogresivnější
- Časté verbální fráze v psané produkci jako znamení nízké úrovně psaného projevu žáka

# Interpretace

- Žákovská argumentace se blíží argumentačnímu registru
- Vyšší podíl na textové variabilitě má registr než idiolekt autora (Cvrček, Laubeová, et al., 2020a)



# Výhled do budoucna

---

- Vliv jazykové úrovně na repertoár lingvistických rysů
- Vliv mateřštiny na jazykovou variabilitu
- Rozdíly mezi rodilými a nerodilými texty v rámci stejných zadání

# Závěrem

---

- První pokus o empirickou evidenci češtiny nerodilých mluvčích a její registrovou variabilitu
- Užitečnost MD modelu v analýze žákovských textů produkováných žáky na podprahové úrovni
- *Non-nativeness* (čes. nepřirozenost) v psané produkci není omezena na jediný rys, ale je mnohem komplexnějším jevem

# Děkuji za pozornost!

---

Adrian Jan Zasina

adrian.zasina@ff.cuni.cz



ÚSTAV  
BOHEMISTICKÝCH STUDIÍ  
Filozofická fakulta  
Univerzita Karlova

# Bibliografie

---

- Asención-Delaney, Y., & Collentine, J. (2011). A multidimensional analysis of a written L2 Spanish corpus. *Applied Linguistics*, 32(3), 299–322.
- Biber, D. (1988). *Variation Across Speech and Writing*. Cambridge University Press.
- Biber, D. (1989). A typology of English texts. *Linguistics*, 27(1), 3–43. <https://doi.org/10.1515/ling.1989.27.1.3>
- Biber, D. (2014). Using multi-dimensional analysis to explore cross-linguistic universals of register variation. *Languages in Contrast*, 14(1), 7–34. <https://doi.org/10.1075/lic.14.1.02bib>
- Biber, D., Gray, B., & Staples, S. (2016). Predicting Patterns of Grammatical Complexity Across Language Exam Task Types and Proficiency Levels. *Applied Linguistics*, 37(5), 639–668. <https://doi.org/10.1093/applin/amu059>
- Cvrček, V., Komrsková, Z., Lukeš, D., Poukarová, P., Řehořková, A., & Zasina, A. J. (2018). Variabilita češtiny: Multidimenzionální analýza. *Slovo a slovesnost*, 79(4), 293–321.
- Cvrček, V., Komrsková, Z., Lukeš, D., Poukarová, P., Řehořková, A., & Zasina, A. J. (2021). From extra- to intratextual characteristics: Charting the space of variation in Czech through MDA. *Corpus Linguistics and Linguistic Theory*, 17(2), 351–382. <https://doi.org/10.1515/cllt-2018-0020>
- Cvrček, V., Laubeová, Z., Lukeš, D., Poukarová, P., Řehořková, A., & Zasina, A. J. (2020). *Registry v češtině*. Nakladatelství Lidové noviny.
- Kim, J., & Nam, H. (2019). How do textual features of L2 argumentative essays differ across proficiency levels? A multidimensional cross-sectional study. *Reading and Writing*, 32(9), 2251–2279. <https://doi.org/10.1007/s11145-019-09947-6>
- Qiu, X., Lo, Y. Y., & Teng, X. S. (2022). Development of L2 disciplinary literacy: A multidimensional analysis. *System*, 110, 1–15. <https://doi.org/10.1016/j.system.2022.102907>
- Staples, S., Biber, D., & Reppen, R. (2018). Using Corpus-Based Register Analysis to Explore the Authenticity of High-Stakes Language Exams: A Register Comparison of TOEFL iBT and Disciplinary Writing Tasks. *The Modern Language Journal*, 102(2), 310–332. <https://doi.org/10.1111/modl.12465>
- Weigle, S. C., & Frigal, E. (2015). Linguistic dimensions of impromptu test essays compared with successful student disciplinary writing: Effects of language background, topic, and L2 proficiency. *Journal of English for Academic Purposes*, 18, 25–39. <https://doi.org/10.1016/j.jeap.2015.03.006>
- Yan, X., & Staples, S. (2020). Fitting MD analysis in an argument-based validity framework for writing assessment: Explanation and generalization inferences for the ECPE. *Language Testing*, 37(2), 189–214. <https://doi.org/10.1177/0265532219876226>