

# Úvod do korpusové lingvistiky 2

1

TYPOLOGIE JAZYKOVÝCH KORPUSŮ

# HLEDISKA

2

- ČAS (synchronní / diachronní)
- TEXT (psaný / mluvený / webový)
- JAZYK (jednojazyčný / paralelní / srovnatelný)
- AUTOR (autorský)
- VÝZKUMNÝ CÍL (obecný / specializovaný)

# ČAS (synchronie a diachronie, řešíme obsah i rozsah)

3

- Bude nás zajímat zkoumat jazyk v nějaké konkrétní době?
- Jaké texty máme k dispozici?
- Jak stanovíme časové rozmezí?
- Jak stanovíme obsah a rozsah
- Jak budeme řešit anotace

# Časové hledisko

4

- synchronní (zachycuje texty, které vznikly ve vymezeném kratším časovém úseku a které jsou v něm recipovány, obtíže – generační a další rozdíly - vágní)
- diachronní (zachycuje texty, které vznikly ve vymezeném delším časovém úseku, úsek je volen tak, aby bylo možné zkoumat vývoj jazyka)

# TEXT (především obsahové hledisko)

5

- Psaný nebo mluvený?
- Formální/připravený nebo neformální/nepřipravený?
- Dialogický nebo monologický?
- Styl a žánr?

# AUTOR (obsahové i rozsahové hledisko)

6

- Známý nebo neznámý?
- Přiznaný nebo anonymní?
- Jak zajistit identifikaci a jak anonymitu?
- Které parametry autorství mohou být pro výzkum jazyka relevantní?

# Výzkumný cíl (obsahové i rozsahové hledisko)

7

- Obecný nebo speciální?
- Popis nebo srovnání?
- Soukromý nebo veřejný?
- Jak velká data (ROZSAH) lze/chceme získat

# Pojetí synchronie v KL

8

- <https://wiki.korpus.cz/doku.php/:pojmy:synchronni>
- Je třeba stanovit
- Lze stanovit s ohledem na žánr (publicistika zastarává rychleji než odborné texty a beletrie)
- Lze stanovit s ohledem na autora (např. v beletrii)
- Lze stanovit s ohledem na text (oportunismus nemusí znamenat nedostatek rozvahy)



# Korpusy řady SYN

9

- SYN2000
- SYN2005
- SYN2010
- beletristické texty - autor narozen po roce 1880, dílo bylo vydáno po roce 1945 a je neustále čteno a vydáváno
- odborné texty - dílo bylo vydáno po roce 1989 (platí méně striktně pro korpus [SYN2000](#), který měl za cíl mapovat širší období)
- publicistika - není starší než 5 let (platí méně striktně pro korpus [SYN2000](#), který měl za cíl mapovat širší období)

# Korpusy řady SYN

10

- SYN2015
- SYN2020
- pro beletrii platí strategie 25 + 75, tj. doba **od prvního vydání nepřesahuje 75 let** (přibližně tři žijící generace) a konkrétní **vydání** díla zařazovaného do korpusu není starší **25 let** (zajištění současné recepce),
- u **odborných** textů platí požadavek prvního vydání v posledních **25 letech**,
- hranice synchronie publicistických titulů zůstává nezměněna, tj. text musí být vydán v **období mapovaném** daným korpusem (v případě SYN2015 je to období let 2010 až 2014).

# KL (právo a etika)

11

- Poskytovatelé textů
- Nekomerční využití
- Autorská práva tvůrců korpusů a korpusových nástrojů – citační etika
- Právní zajištění při získávání dat pro korpusy
- Anonymizace citlivých údajů v korpusech

# Mluvené korpusy

12

- pmk/bmk
- řada ORAL
- DIALEKT
- multimodální korpus (grafický přepis, fonetickou původní i průvodní nahrávku a (stále velmi řídké) ještě někdy i nafilMOVANÝ záznam)
- <https://www.clarin.eu/resource-families/spoken-corpora>

# Primát mluveného jazyka a jazykové korpusy

13

- Nedostatek mluvených textů pro diachronní výzkum
- Vysoké náklady na pořízení mluvených textů v porovnání s texty psanými
- Otázka přirozenosti a etická hlediska
- Mluvený jazyk a žánr
- Psaný jazyk a žánr
- Vícejazyčnost v KL

# SEU

[https://en.wikipedia.org/wiki/Survey\\_of\\_English\\_Usage](https://en.wikipedia.org/wiki/Survey_of_English_Usage)

14



# Typy korpusů ČNK

15

[Všechny korpusy](#)

řada SYN řada ORAL InterCorp synchronní diachronní mluvený psaný webový současná verze starší verze  
čeština cizojazyčný paralelní srovnatelný autorský reprezentativní akviziční specializovaný

# Příklad otázek v testu

16

- Charakterizujte pojetí synchronního korpusu řady SYN
- K čemu mohou sloužit autorské korpusy?
- Uveďte důvody malého rozsahu korpusů mluveného jazyka?
- Existují korpusy dialektologické?
- Je možné zkoumat synchronní stav jazyka starších jazykových období?
- Je možno zohlednit časové hledisko u specializovaných korpusů (vývoj jazyka jedince, vývoj podoby žánru)?