

Úvod do korpusové lingvistiky 2

1

TYPOLOGIE JAZYKOVÝCH KORPUSŮ

HLEDISKA

2

- ČAS
- TEXT
- JAZYK
- AUTOR
- VÝZKUMNÝ CÍL

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

TEXT (především obsahové hledisko)

4

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

AUTOR (obsahové i rozsahové hledisko)

5

- 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)

6

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

Pojetí synchronie v KL

7

- <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)

Primát mluveného jazyka a jazykové korpusy

8

- 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

Autor

9

- Anonymita v diachronii
- Ochrana autora a získávání dat pro korpusy
- Sociolingvistická metadata - co vše je třeba brát v úvahu
- Hraje překladatel roli v KL?

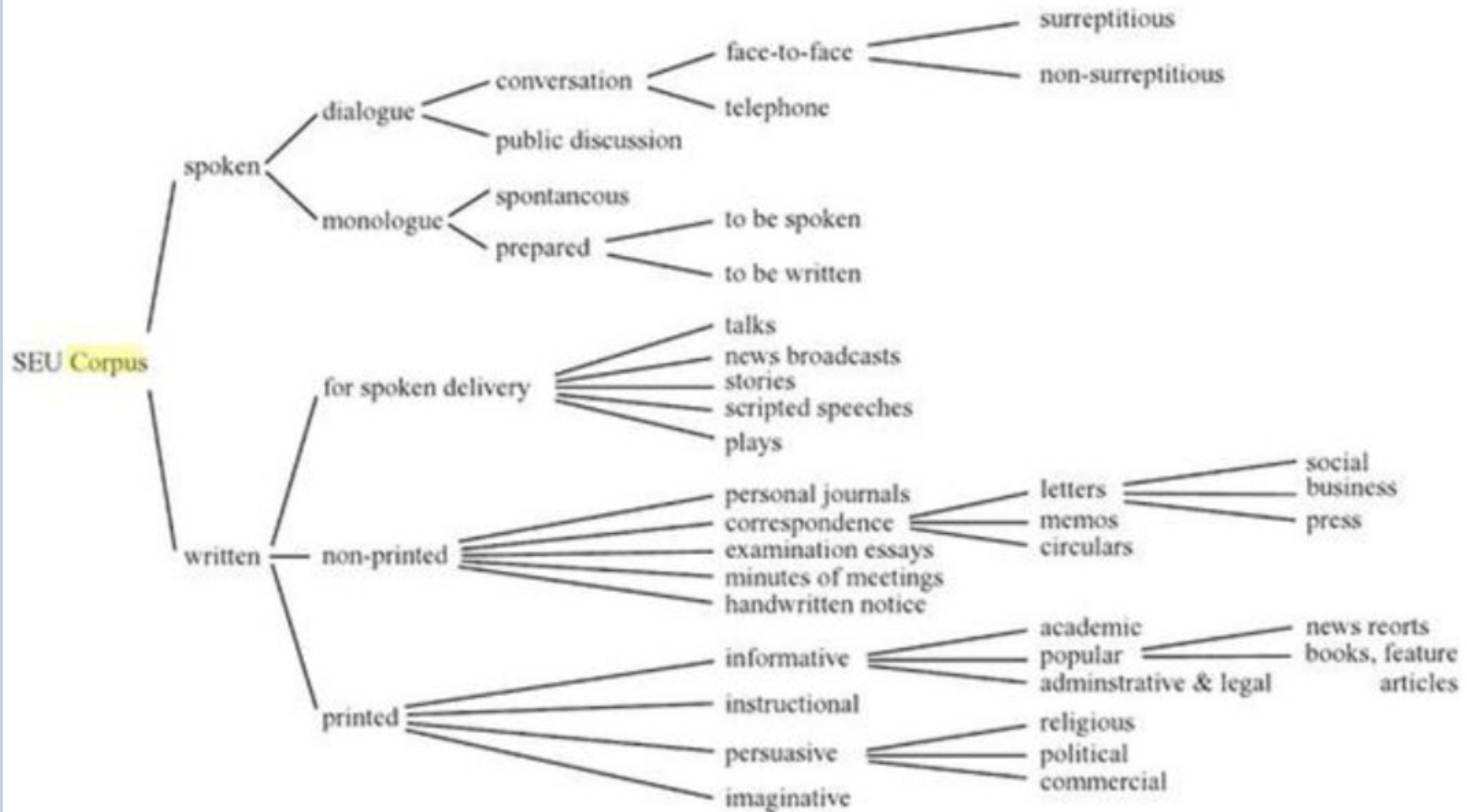
Výzkumný cíl (jazyk a nejen on)

10

- Podpora velkých projektů
- Národní a internacionální podpora (snahy o sjednocení a sdílení)
- Elektronizace a korpusové zpracování nejrůznějších archívů jak jedna z výzev KL (digital humanities)

SEU

11



- <https://www.english-corpora.org/>
- <https://www.youtube.com/watch?v=NGxvJU7iS2I>

Příklad otázek v testu

13

- 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?