

+Mgr. Vladislav Belev: Analysis of power dynamics through speech acts in the video game Detroit: Become Human and Mass Effect series

Mgr. Bc. Jana Hallová: Multimodální kreativita na sociálních médiích

<https://www.clarin.eu/resource-families/multimodal-corpora>

Multimodal corpora

Introduction

Multimodal corpora are data collections used to study how two or more modalities interface with one another in human communication. In this sense, multimodal corpora are often collections of video and speech recordings accompanied with transcriptions and gesture annotations, although multimodal corpora of textual data supplemented with images exist as well. Such corpora can be used for “the exploration of a range of lexical, prosodic and gestural features of conversation, and for investigations of the ways in which these features interact in real, everyday speech (Abuczki and Baiat Ghazaleh [2013](#): 88).

The CLARIN infrastructure offers 17 multimodal corpora, 13 of which are monolingual (English, Estonian, Finnish, French, German, Greek, Hungarian, Italian, Slovenian, and Zulu). These corpora are richly annotated for various verbal and non-verbal elements of communication, such as body gesture, gaze direction, and head, eye, and lip movement.

For comments, changes of the existing content or inclusion of new corpora, send us an email.

Akviziční korpusy

Akviziční korpusy představují speciální typ jazykových [korpusů](#), které slouží primárně pro studium procesů spojených s osvojováním řeči a s jejím vyučováním jako prvního i druhého jazyka. Poskytuje informace o užívání jazyka mluvčími (různého věku), kteří si ho dosud neosvojili na úrovni dospělého rodilého mluvčího.

Akviziční korpusy mohou mít podobu [psaného korpusu](#) i [korpusu mluveného](#), jsou zaměřené na rodilé i nerodilé mluvčí (viz [žákovské korpusy](#)). Specifickým rysem některých akvizičních korpusů je tzv. **chybová anotace** naznačující, v jakých místech se mluvčí nebo pisatel odchýlil od běžného úzu nebo normy. Tyto odchylky jsou v rámci systému anotace klasifikovány a je podle nich pak možné vyhledávat.

V rámci ČNK jsou k dispozici korpusy [SCHOLA2010](#), [CzeSL-plain](#) a [CzeSL-SGT](#).

<http://akces.ff.cuni.cz/>
<http://akces.ff.cuni.cz/node/155>

Žákovský korpus

Žákovské korpusy (learner corpora) obsahují texty většinou nerodilých mluvčích, kteří se jazyk teprve učí. Povětšinou se jedná o [psané](#) (školní) práce, zřídka nacházíme [mluvené](#) žákovské korpusy.

Význam žákovských korpusů vzrůstá nejen s jejich rozsahem ale také s jejich [anotací](#). Pokud je součástí jejich metadat také tzv. chybová anotace, je možné klasifikovat jednotlivé odchylky od běžného úzu, zjišťovat jejich četnost, příp. i motivaci. Součástí metadat bývá také rozšířená informace o pisateli, která je často obohacena o relevantní údaje vztahující se k jeho věku, mateřskému jazyku, úrovni jeho znalostí cílového jazyka apod.

Žákovské korpusy zachycují jazyk mluvčích/pisatelů, kteří dosud nezvládají cílový jazyk dokonale; v tomto smyslu je možné je chápat jako specifický případ [korpusů akvizičních](#). V rámci ČNK jsou k dispozici žákovské korpusy [CzeSL-plain](#) a [CzeSL-SGT](#).

Mgr. Alžběta Bohme Ryšková: E studio de caso de bilinguismo temprano

Co se týče korpusů, měla byste se podívat na články zaměřené na tzv. akviziční korpusy.

Mgr. Kristýna Lorenzová: Lo sviluppo dei marcatori discorsivi nelle interlingue di apprendenti cechi e slovacchi L2: uno studio longitudinale

Žákovské korpusy

<https://www.sketchengine.eu/merlin-learner-corpus/>

<https://www.merlin-platform.eu/>

<https://www.youtube.com/watch?v=aaECpYdeqaQ>

Mgr. Yurii Chybras: Výslovnost a identita: Fonologie a vnímání v angličtině se slovanským přízvukem

Mgr. Pavel Čanecký: Vztah mezi percepcí a produkcí fonologického inventáře univerzitních studentů anglického jazyka s využitím digitální technologie ve výuce

Mgr. Laurent Georges Canal: La notion d'argot dans les corpus électroniques et dans les réactions spontanées des locuteurs francophones

Domnívám se, že buď nějaký korpus, z něhož byste mohl čerpat, chcete vytvořit, nebo že víte, že nějaký specializovaný korpus vytvořen je. Ve druhém případě bych si dovolila poznamenat, že z korpusů, které jsou přístupné volně, lze pomocí korpusových nástrojů pomocí funkce "vytvoř subkorpus"/create subcorpus/ vytvořit vhodný korpus pro konkrétní potřebu, a sice pomocí metadat (žánr,).

Mgr. Matyáš Foltýn: Analýza rozsahu vlivu liturgických a okolních jazyků na lexikum židovských diasporních jazyků

Tohle je velmi zajímavé téma. Korpusy se pro výzkum lexika používali. Liturgické jazyky jsou zase jazyky psaných památek, takže z hlediska KL není tak obtížné je vytvořit (nechybějí data). Horší je to s nástroji, tam je třeba překonat obtíže, které souvisejí s převedením textů do počítačově čitelné podoby (transkripce, transliterace, ... různocnění, ... až po to, že by badatel měl mít informace o nástrojích OCR a dalších, které podporují proces převedení textu do počítačově čitelného formátu. Také, pokud jde o nelatinkové texty, třeba i vědomost o tom, jaké jsou možnosti - iniciativa TEI).

Mgr. Emma Jackovičová: Functional sentence perspective theory: cross-linguistic analysis of information structure and its perception

Měla byste si promyslet: a) existují korpusy (syntakticky značkové treebanky - PDT - <https://ufal.mff.cuni.cz/courses/npl082>), v nichž lze hledat struktury, které odhalí vlastnosti funkční větné perspektivy a b) víte, jak se v nich hledá/jak se kladou dotazy.

Mgr. Martin Juřen: Česká abstraktní nepočitatelná substantiva

Českých korpusů je k dispozici dost. Data máte. Vzhledem k tomu, že nemůžete využít značkování vlastnosti počitatelnost/nepočitatelnost, tak budete muset jít jednak po klíčových slovech, jednak po lexikálních jednotkách v kolokačních profilech klíčových nepočitatelných substantiv.

Mgr. Kristýna Mottlová: Postavení humoru v barokní homiletické tvorbě v kontextu historického modelu vývoje komična

Zaměřila bych se na využití subkorpusu. Ten lze sestavit pomocí nástroje kontext přes Vokabulář webový, středočeskou banku, vyhledat texty, které jsou v metadatech označeny jako "kázání" a ty pak dále podle potřeby filtrovat.

Funkce:

Korpus/Vytvoř subkorpus

zadat v metadatech kázání

https://korpus.vokabular.ujc.cas.cz/first_form?corpname=SSTB-1.1.23.1&usesubcorp=KAZANI

V Diachronním odd. ÚJČ se lze obrátit na našeho absolventa PLIN Mgr. Ondřeje Svobodu (e-mail: svoboda@ujc.cas.cz).

Mgr. Lucia Šalagová: Kontrastivní analýza slovesných časů v litevštině a finštině se zřetelem k analytickým formám časů minulých (perfektum a plusquamperfektum)

Pokud budete pracovat s korpusy, pak pro vyhledávání bude Vaším problémem naučit se pracovat s tagy, přes které budete chtít hledat analytické formy. Mám s tím zkušenosti z českých korpusů. V jazycích s jednoduššími pravidly slovosledu je práce na sestavení dotazu jednodušší než u jazyků s pravidly složitějšími. Takže lingvisticky bych vyšla z této startovní čáry a zkoumala bych, jak se správně ptát (cql jazyk).

Mgr. et Mgr. Adéla Kereková: Generační posuny v běžné mluvě na Uničovsku

Mgr. Šimon Temer: Jazyk Čechů v srbském Banátu

Disertační projekt je velmi obecně formulovaný. Ze zkušenosti víme, že studentovi nejvíc pomůže, když se zaměří na jevy, které lze uchopit. Korpusová lingvistika je empirická disciplína. Pokud jste dosud nezvolil jevy, na které se chcete zaměřit, mějte na paměti, abyste je volil se zřetelem k tomu, a) jak rychle (časový horizont napsání disertace) získáte data (existují data / musíte je získat), b) jste-li schopen (vzhledem k jevu) mít dost dat (kvantita dat má vliv na kvalitu výzkumu) na to, abyste mohl něco vyzkoumat. Až pak je možné uvažovat o korpusovém výzkumu.

Bartłomiej Wieczorek: MUF

Hugo Kuhl: DU/FR