

# Úvod do korpusové lingvistiky

10



**ORGANIZACE, ČASOPISY, PUBLIKACE,  
KONFERENCE**

# Korpusová lingvistika



- Korpusová lingvistika
- Počítačová lingvistika (Computational Linguistics)
  - NLP (Nature Language Processing)
  - Digital Humanities
- Mezinárodní i národní konference
- Odborné časopisy

# TEI



- TEI (Text Encoding Initiative) od r. 1987
- důležitá nezisková organizace akademických pracovišť, výzkumných projektů a odborníků
- vyvíjí a udržuje standardy pro reprezentaci textu v digitální podobě
- standardy jsou závazné mj. pro tvorbu korpusů ve společenských vědách a lingvistice
- od r. 1994 řídí značkování jazyka a struktury textu (XML, HTML)

# Mezinárodní organizace, konference, časopisy



- ACL (Association for Computational Linguistics)
  - založena 1962 (Association for Machine Translation and Computational Linguistics AMTCL, od r. 1968 ACL)
- časopis Computational Linguistics (MIT Press)
  - prestižní časopis počítačové lingvistiky
  - vychází čtvrtletně od r. 1988, od r. 2009 Open Access
- konference ACL od r. 2002, prestižní konference
  - 2021 Bangkok, Thajsko
  - 2017 Vancouver, Kanada
  - 2016 Berlín, Německo
  - 2015 Peking, Čína
  - 2007 Praha (možnost účasti českých počítačových lingvistů a informatiků i bez příspěvků, jinak není snadné se na konferenci dostat)

# Mezinárodní organizace, konference, časopisy



- ICAME (International Computer Archive of Modern and Medieval English)
  - mezinárodní organizace lingvistů a informatiků pracujících s elektronickými texty v angličtině
  - University of Bergen, Norsko, založena v Oslu 1977
- ICAME conference
  - 2021 42. roč., Dortmund, Německo
  - 2017 38. roč., Praha, ČR
  - 2016 37. roč., Hong Kong
  - 1979 1. roč., Bergen, Norsko
- ICAME Journal (od 1978) – odborný časopis, minimálně jednou ročně

# Mezinárodní organizace, konference, časopisy



- ACH (Association for Computers and the Humanities)
  - založena 1978
  - od 2002 součástí ADHO (Alliance of Digital Humanities Organizations)
- konference Digital Humanities conference, od 1989
- časopis Computers and the Humanities od 2004
- od 2005 Language Resources and Evaluation
- časopis Literary and Linguistic Computing (Oxford University Press)

# Další mezinárodní časopisy



- International Journal of Corpus Linguistics (John Benjamins Publishing Company, Amsterdam)
  - v redakční radě byl František Čermák
  - od r. 1996, čtvrtletně od 2005, e-Journal, možnost přihlášení přes Institutional Login (MU)
- Corpora (Edinburgh University Press)
  - Lancaster University
  - od r. 2006, 2krát ročně

# Konference (ČR)



- Korpusová lingvistika Praha, pořádá ÚČNK, z konferencí vyšly sborníky s příspěvky Praha 2011 (3 sborníky)
  - 1 InterCorp
  - 2 Výzkum a výstavba korpusů
  - 3 Gramatika a značkování korpusů
- Praha 2014 – 20 let mapování češtiny
  - k 20letému výročí ÚČNK
- Praha 2016 – Od mluvené češtiny k psané
- SlaviCorp 2018, ÚČNK, Praha, mezinárodn



# Konference v ČR



- Gramatika a korpus – mezinárodní konference, ÚJČ AV ČR
  - 1. Praha, 2005
  - 2. Liblice u Mělníka, 2007
  - 3. Mannheim, 2009 – Grammar & Corpora
  - 4. Praha, 2012 (100. výročí narození Miloše Dokulila)
  - 5. Varšava, 2014 – Grammar & Corpora
  - 6. Mannheim, 2016 – Grammar & Corpora
  - 7. Paříž, 2018 – Grammar & Corpora
  - 8. Krakow, 2020 – Grammar & Corpora
- Korpus jako zdroj dat o češtině, Šlapanice u Brna, 2004
  - sborník Karlík, P. (ed.): Korpus jako zdroj dat o češtině. Brno: MU, 2004.
- Čeština v mluveném korpusu, Praha, 2007 (věnovaná pouze mluveným korpusům)
- InterCorp, Praha, 2009 (věnovaná pouze paralelnímu korpusu)

# Konference v ČR/SR



- TSD (Text, Speech and Dialogue) počítačové zpracování textu, řeči a dialogu - počítačová, ale i korpusová lingvistika
  - 24. ročník – Olomouc
  - Brno (FI MU)
  - Plzeň (FAV ZČU)
- na Slovensku **Slovko** (Slovenský národný korpus, Jazykovedný ústav Ľudovíta Štúra)
  - mezinárodní, jednou za dva roky

# KGA



- Korpus – gramatika – axiologie
  - od r. 2010, 2krát ročně
- Ústav pro jazyk český AV ČR
  - v redakční radě za Brno – Klára Osolsobě, Karel Pala
- <https://nlp.fi.muni.cz/kga/>
- [Archiv - Ústav pro jazyk český](#)

## ***KORPUS – GRAMATIKA – AXIOLOGIE***

Časopis pro korpusový výzkum a hodnocení jazyka  
Journal for corpus research and evaluation of language

Poslání časopisu / Mission of the journal  
objednávka / order



- sborníky odborných článků
- příručky pro práci s korpusem
- mluvnické založené na korpusu
- slovníky založené na korpusu

# Publikace



- Šulc, M. Korpusová lingvistika: první vstup. Praha, UK, 1999.
- Čermák, F., Klímová, J. a Petkevič, V. (eds.). Studie z korpusové lingvistiky. Praha: Karolinum, 2000.
- Kocek, J., Kopřivová, M., Kučera, K. (eds.). Český národní korpus – úvod a příručka uživatele. Praha: ÚČNK FF UK, 2000.
- Čermák, F., Blatná, R. (eds.). Jak využívat Český národní korpus. Praha: NLN, 2005.
- Čermák, F. Korpus a korpusová lingvistika. Praha: Karolinum, 2017.
- Přehled publikací z produkce ÚČNK

<https://ucnk.ff.cuni.cz/cs/veda-a-vyzkum/publikace/>

# Gramatiky



- Štícha, F. (ed.): Možnosti a meze české gramatiky. Praha: Academia, 2006.
- Štícha, F. (ed.): Kapitoly z české gramatiky. Praha: Academia, 2011.
- Štícha, F. (ed.): Akademická gramatika spisovné češtiny. Praha: Academia, 2013.
- Cvrček, V. a kol.: Mluvnice současné češtiny 1. Jak se píše a jak se mluví. Praha: Karolinum, 2010.
- Panevová, J. a kol.: Mluvnice současné češtiny 2. Syntax češtiny na základě anotovaného korpusu. Praha: Karolinum, 2014.

# Slovníky



- menší speciálně zaměřené slovníky (viz odkaz na web ÚČNK, seznam publikací)
  - Čermák, F. – Křen, M. (eds): Frekvenční slovník češtiny. Nakladatelství Lidové noviny, Praha 2004.
    - založen na opraveném korpusu SYN2000 – FSC2000
  - stále připravovaný a postupně zveřejňovaný
- Akademický slovník současné češtiny
  - ÚJČ AV ČR, oddělení současné lexikografie a lexikologie
  - zveřejněna písmena A, B, C, Č
  - využívá výsledky korpusové lingvistiky a počítačové lexikografie

# CLARIN <https://korpus.cz/clarin>



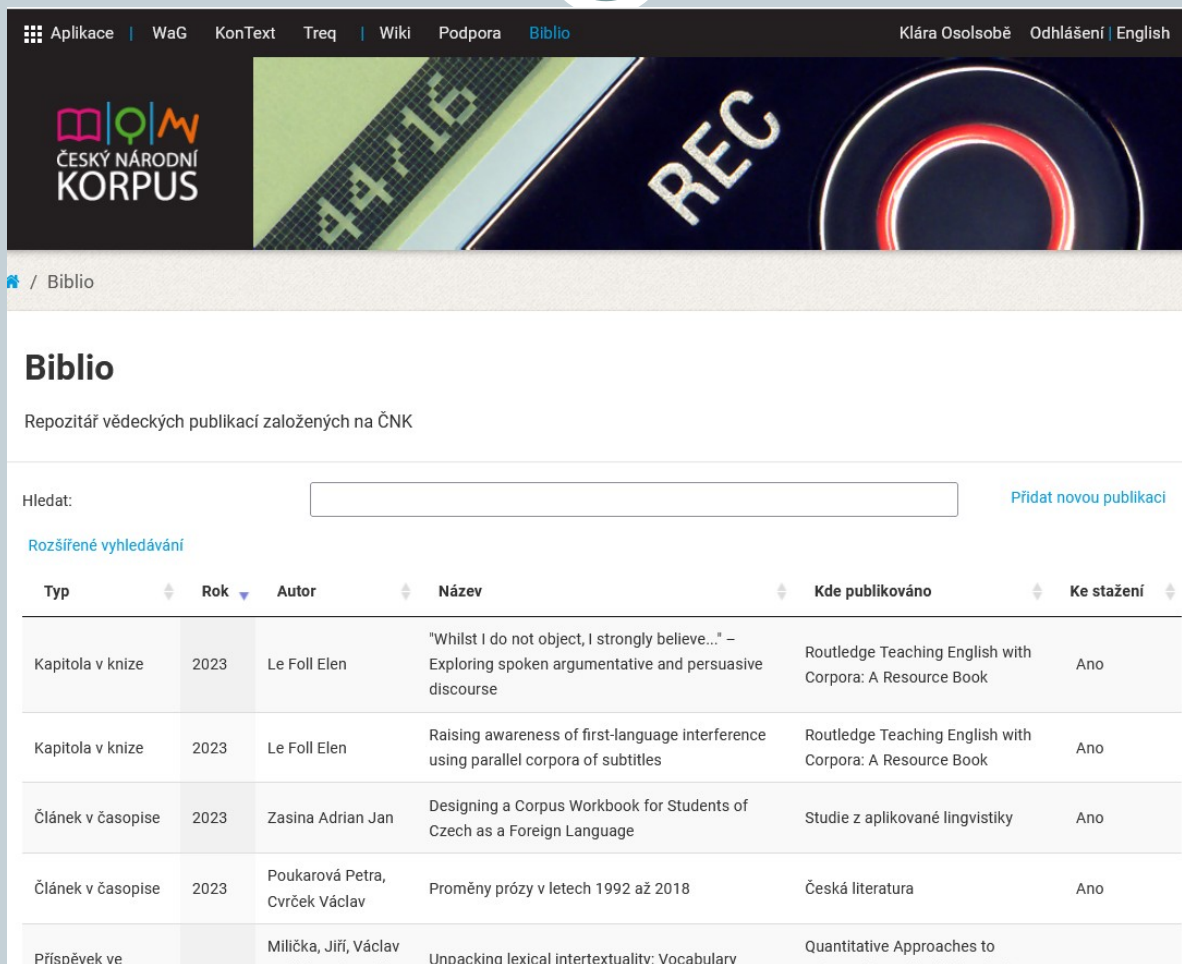
## Podpora a zdroje informací

Wiki	K-centrum projektu ČNK poskytuje informace, konzultace a technickou podporu v oblasti korpusové lingvistiky se zaměřením na empirický výzkum češtiny. Je součástí <a href="#">K-center</a> evropské infrastruktury <a href="#">CLARIN</a> zaměřené na jazykové zdroje a nástroje pro práci s nimi pro humanitní a společenské vědy.
Podpora	
Biblio	
Rada ČNK	
Jazyková data	
Pro školy	
K-centrum CLARIN	

[K-centrum CLARIN](#) →



# Biblio <https://www.korpus.cz/biblio/appeal>



The screenshot shows the Biblio website interface. At the top, there is a navigation bar with links for 'Aplikace', 'WaG', 'KonText', 'Treq', 'Wiki', 'Podpora', and 'Biblio'. On the right side of the navigation bar, it says 'Klára Osolsobě', 'Odhlášení', and 'English'. Below the navigation bar is a banner image with the logo 'ČESKÝ NÁRODNÍ KORPUS' and the word 'REC'. The main content area has a breadcrumb trail ' / Biblio' and a title 'Biblio' with the subtitle 'Repozitář vědeckých publikací založených na ČNK'. There is a search bar labeled 'Hledat:' and a button 'Přidat novou publikaci'. Below the search bar is a link 'Rozšířené vyhledávání'. The main content is a table of publications with columns: Typ, Rok, Autor, Název, Kde publikováno, and Ke stažení.

Typ	Rok	Autor	Název	Kde publikováno	Ke stažení
Kapitola v knize	2023	Le Foll Elen	"Whilst I do not object, I strongly believe..." – Exploring spoken argumentative and persuasive discourse	Routledge Teaching English with Corpora: A Resource Book	Ano
Kapitola v knize	2023	Le Foll Elen	Raising awareness of first-language interference using parallel corpora of subtitles	Routledge Teaching English with Corpora: A Resource Book	Ano
Článek v časopise	2023	Zasina Adrian Jan	Designing a Corpus Workbook for Students of Czech as a Foreign Language	Studie z aplikované lingvistiky	Ano
Článek v časopise	2023	Poukarová Petra, Cvrček Václav	Proměny prózy v letech 1992 až 2018	Česká literatura	Ano
Príspevek ve	2023	Milička, Jiří, Václav	Unpacking lexical intertextuality: Vocabulary	Quantitative Approaches to	

# Příklady otázek v testu



- Vyjmenuj důležité konference specializované na KL
- Vyjmenuj alespoň dvě gramatiky založené na korpusech
- Vyjmenuj důležité časopisy zaměřené na KL