

PLIN041 Vývoj počítačové lingvistiky

Korpusová lingvistika v ČR Počítačová lingvistika v ČR – Brno Slovensko

Mgr. Dana Hlaváčková, Ph.D.

Korpusová lingvistika v ČR

- **lexikografické počátky**
- 1988 *Iniciativní skupina pro přípravu počítačových korpusů, textů a slovníků* (Pala, Čermák, Schmiedtová, Hajičová ad.)
 - záštita [Kybernetické společnosti](#), později Jazykovědného sdružení ČSAV
- *Skupina pro počítačový fond češtiny* – Čermák, Králík, Pala, Hajič, Hajičová, Sgall, Schmiedtová, Benko, Kučera
- Čermák, Králík, Pala – *Počítačová lexikografie a čeština* (počítačový fond češtiny), SaS, ročník 53, č. 1, 1992.
- **1993–95 Počítačový korpus českých psaných textů** (GA ČR)
- spojení s *Nakladatelstvím Lidové noviny* – podpora pro vydání nového výkladového slovníku
- 2023 – *Akademický slovník současné češtiny (písmeno G)*

Korpusová lingvistika v ČR

- 1994 – založení **Ústavu Českého národního korpusu** a počátek projektu **Český národní korpus**
- ředitel ÚČNK – prof. František Čermák (do 2013)
- 1995 – cesta do Velké Británie po centrech korpusové lingvistiky – Pala, Čermák, Petkevič, Schmiedtová
 - Oxford University Press, University of Oxford – **Patrick Hanks**
 - School of English, Birmingham City University – **John Sinclair**
 - Lancaster University – **Geoffrey Leech**

Korpusová lingvistika v ČR

- počátky pracoviště obtížné (sklep FF UK)
 - nedostatek počítačů
 - nedostatek pracovníků
 - odpor některých lingvistů k novému směru výzkumu (ředitel ÚJČ F. Daneš)
- postupně rozrůstání týmu – lingvisté, matematici, programátoři
- budování prvního korpusu SYN2000
- problematika tokenizace, lemmatizace a tagování
- spolupráce s ÚFAL MFF UK, ÚTKL FF UK, FF a FI MU

Korpusová lingvistika v ČR

- příprava **korpusového manažeru**
 - nepodařilo se získat ze zahraničí
 - Pavel Rychlý – **CQP** (*Corpus Query Processor*), využívá jazyk CQL
 - Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung, prof. Ulrich Heid, autoři CQP Schulze a Christ
- **Manatee – Bonito** – Pavel Rychlý (dizertační práce, korpusový manažer *klient – server*)
- webové rozhraní **Bonito2, Sketch Engine, NoSketch Engine, KonText**

Korpusová lingvistika v SR

- počátky – **Jozef Mistrík**, frekvenční slovník (1969)
 - materiál v obdobném složení jako Brown Corpus (zcela nezávisle)
 - statistické výpočty na počítači (n. p. Slovnaft)
- 1988–1994 **Emil Páleš** – formální model flexe slovenštiny (2 tis. lex. jednotek)
 - později rozšířen na cca 40 tis., **Benko**, Hašanová, Kostolanský (model *BHK*, Krátky slovník slovenského jazyka)
- 1991–1993 *Spoločná pracovná skupina pre počítačovú lingvistiku* (JÚĽŠ + Informačné centrum SAV)
 - grant *Metodika budovania korpusu textov a bázy dát slovenského jazyka*
- 2002 vznik *Oddelenia Slovenského národného korpusu* (JÚĽŠ SAV)
- prví korpus r. 2006

Filozofická fakulta UJEP/MU Brno

- bohemistika na FF UJEP se od roku 1921 postupně odděluje od slavistiky, české literatury a obecné jazykovědy
- od poč. 90. let **Ústav českého jazyka**
- pracoviště se soustřeďuje hlavně na:
 - diachronní studie a dialektologii (Trávníček, Havránek, Kellner, Lamprecht, Šlosar)
 - syntax (Bauer, Grepl, Karlík)

Filozofická fakulta UJEP/MU Brno

- **Karel Pala** (1939 Zlín – 2023 Brno)
- 1956–1960 Vysoká škola jazyka ruského a literatury v Praze, specializace překladatelská (čeština a ruština), rektor B. Havránek, později FF UK v Praze
 - vliv Pavla Materny a Pavla Tichého (TIL)
- 1962–1964 postgraduální kurz z matematické lingvistiky, logiky a informatiky, UK a ÚJČ (pod vedením prof. P. Sgalla)
- 1964 nástup na FF UJEP Brno
- 1967 získání titulu PhDr. FF UJEP Brno
- 1972–1973: lektor českého jazyka na School of Slavonic and East European Studies, UCL, předčasně ukončeno pro odmítnutí spolupráce s StB

Filozofická fakulta UJEP/MU Brno

- 1973 získání CSc. v oboru český jazyk, Univerzita Karlova, školitel prof. P. Sgall
- 1993 docentská habilitace v oboru český jazyk (se zaměřením na počítačnou lingvistiku a počítačové zpracování přirozeného jazyka), MU Brno
- 1993–1995: lektor českého jazyka na School of Slavonic and East European Studies, University of London
- **1964–1995 FF UJEP/MU** (přesun z Prahy do Brna na základě konkurzu na místo odb. asistenta, přijímal ho děkan prof. Milan Jelínek
- **od 1995 FI MU** – Centrum zpracování přirozeného jazyka

Filozofická fakulta UJEP/MU Brno

- *Lidová píseň a samočinný počítač*, 1976, Pala, Holý, Štědroň
- *Česká morfologie a syntax v PROLOGU*, 1987, Pala, Osolsobě, Franc
- *Logická analýza přirozeného jazyka*, 1989, Materna, Pala, Zlatuška
- *Základy výpočetní techniky pro filology*, 1989, Pala, Halasová
- *Základy počítačové lingvistiky*, 1992, Pala, Osolsobě

Filozofická fakulta UJEP/MU Brno

- **Karel Pala** od 1964 *Katedra českého jazyka, slovanské a obecné jazykovědy* Filozofické fakulty UJEP
- prosadil užívání počítačů na katedře
- **Seminář počítačové lingvistiky (1992–1995)**
 - spolupráce s logikem Pavlem Maternou, informatikem Jiřím Zlatuškou, anglistou Alešem Svobodou
 - morfologická a syntaktická analýza
- Karel Pala, **Klára Osolobě** a Stanislav Franc
 - morfologicko-syntaktický analyzátor **klara** využívající jazyk Prolog a aparát DC gramatik (Definite Clause Grammars)

Filozofická fakulta UJEP/MU Brno

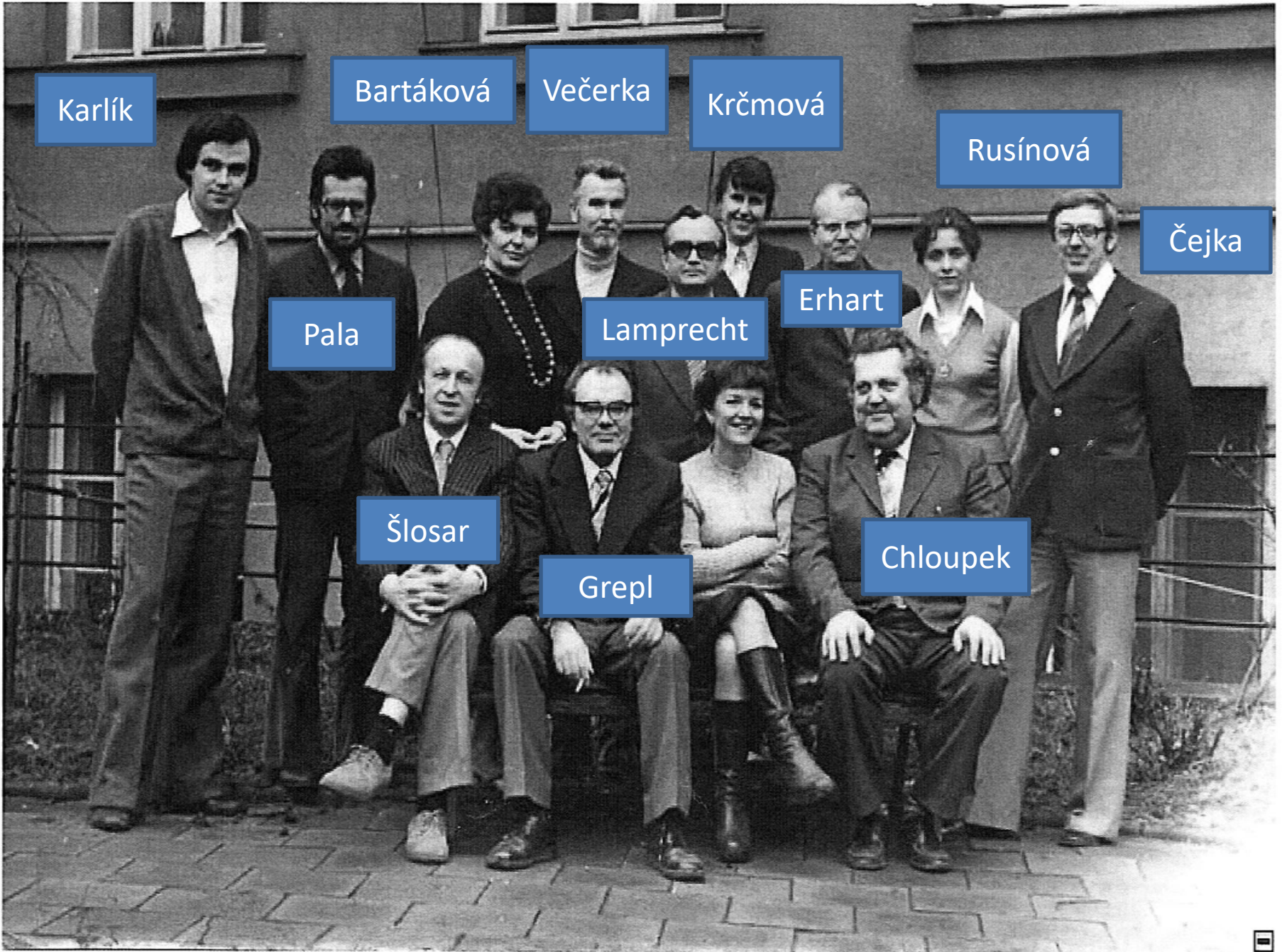
- **jazykový korektor** pro textový editor **T602** (Osolsobě, Franc)
- morfologický analyzátor **Xantypa** (Osolsobě, Franc)
- morfologický analyzátor **LEMMA** – (Osolsobě, Pavel Ševeček)
- mluvený korpus **BMK** (Hladká, 1994–1999, zveřejněn 2002)

Centrum zpracování přirozeného jazyka

- korpusový manažer (Bonito), korpusové nástroje
- morfologická analýza (Ajka) a desambiguace (desamb)
- syntaktická analýza (synt)
- počítačová lexikografie (gedit)
- slovotvorná analýza (Deriv)
- Czech WordNet
- slovesná valence (Brief) a sémantika

Počítačová lingvistika na FF MU – současnost

- 2010 vznik studijního oboru **Český jazyk se specializací počítačová lingvistika**, od 2019 studijní program **Počítačová lingvistika**
- 2012–2014 OP VK *Mezi bohemistikou a informatikou*
- rozvoj spolupráce s řadou brněnských firem
- absolventi PLINu v **doktorském studiu**, spolupráce při výuce PLINu
- **projekty TA ČR**
 - Adaptabilní mobilní aplikace k výuce češtiny pro cizince, CzechME
 - Webový pravopisný, gramatický a typografický korektor pro český jazyk, Opravidlo



Karlík

Bartáková

Večerka

Krčmová

Rusínová

Čejka

Pala

Lamprecht

Erhart

Šlosar

Grepl

Chloupek