

PLIN041 Vývoj počítačové lingvistiky  
Počítačová lingvistika v ČR – Praha

Mgr. Dana Hlaváčková, Ph.D.

# Pracoviště v ČR – vývoj

- **Ústav pro jazyk český ČSAV, Praha**
  - *Oddělení matematické a aplikované lingvistiky (1961–1985)*
  - *Mechanografická laboratoř (1967–1975)*
- **Filozofická fakulta UK, Praha**
  - *Oddělení pro teorii strojového překladu (1957– 1962)*
  - *Laboratoř algebraické lingvistiky (1968–1972)*
  - **Ústav teoretické a počítačové lingvistiky (1990)**
  - **Ústav Českého národního korpusu (1994)**
- **Matematicko-fyzikální fakulta UK, Praha**
  - **Ústav formální a aplikované lingvistiky (1990)**
- **Filozofická fakulta UJEP/MU, Brno**
  - **Ústav českého jazyka (bohemistika od 1921, poč. 90. let ÚČJ, 2010 PLIN)**
- **Fakulta informatiky MU, Brno (1994)**
  - **Centrum zpracování přirozeného jazyka (1995)**

# Ústav pro jazyk český

- **Oddělení matematické a aplikované lingvistiky ÚJČ ČSAV**
- založil 1961 **Lubomír Doležel** (1961–65), (nar. 1922), lingvista a literární teoretik (stylistika, naratologie, teorie fikce), od 1968 Kanada, University of Toronto, spoluzakladatel teorie fikčních světů, od r. 2009 zpět v ČR
- **Marie Těšitelová** – vedoucí pracoviště (1965–1985)

# Ústav pro jazyk český

- **přípravné období**
  - překlady děl z oblasti matematické lingvistiky
  - semináře a referáty
  - vzdělávání lingvistů v matematických metodách (teorie pravděpodobnosti a statistika, Matematický ústav ČSAV)
  - spolupráce matematiků a logiků
- **výzkum**
  - kvantitativní lexikální a gramatické analýzy psané i mluvené češtiny
- sborník překladů ***Teorie informace a jazykověda***, 1964
- ***Cesty moderní jazykovědy***, 1964, část o kvantitativní lingvistice

# Marie Těšitelová (1921–2011)

- studovala bohemistiku a historii na FF UK až po r. 1945
- už během studia se dostává ke kvantitativní lingvistice
- *O frekvenci slov a tvarů v Čapkově románu Život a dílo skladatele Foltýna* (Naše řeč 32, 1948, pod vedením V. Šmilauera)
- Výzkumný ústav pedagogický – příprava učebnic češtiny
- s J. V. Bečkou a J. Jelínkem práce na frekvenčním slovníku
- od 1956 v ÚJČ – podílí se na SSJČ, výkonná redaktorka SaS

# Marie Těšitelová

- 1965–1985 vede oddělení Matematické a aplikované lingvistiky
- 1990 odchází do penze
- žila osaměle, její láskou byla kvantitativní lingvistika, přestože používání počítače pro ni nebyl problém, dopisy psala vždy rukou
- nazývána *první dámou české lingvistiky*

# Marie Těšitelová

- *Korpus věcného stylu* (1971–1985), ručně morfologicky a syntakticky značkový
- 540 tis. slov, každý text 3 000 slov
- publicistické, odborné a administrativní texty (psané i mluvené)
- texty přepsány na stroji, ruční anotace, vyděrováno na děrné štítky
- sálové počítače Tesla 200 a IBM 370 (Ústav teorie informace a automatizace ČSAV)
- plánovala se další půlmilionová část z beletrie
- nakladatelství Academia odmítlo vydání
- *Český akademický korpus*, kompatibilní s PDT, ÚFAL MFF UK, 2007

# Marie Těšitelová

- příspěvky v časopisech a sbornících publikovaných doma i v zahraničí, články psala v němčině i angličtině (v té době neobvyklé)
- *O morfologické homonymii v češtině* (1966)
- *Otázky lexikální statistiky* (1974)
- *Slovník slovanské lingvistické terminologie*, 1977, 14 jazyků, ved. A. Jedlička
- vysokoškolská skripta *Kvantitativní lingvistika* (1977, přepr. 1987)
- *Využití statistických metod v gramatice* (1980)



# Marie Těšitelová

- značný autorský podíl na knížce *O češtině v číslech* (1987)
- série studií o kvantitativních vlastnostech textů různých stylů na základě *Korpusu věcného stylu*
- spolu s J. Petrem a J. Králíkem je autorkou frekvenčního *Retrográdního slovníku současné češtiny* (1986), na základě *Korpusu věcného stylu*
- *Quantitative Linguistics* (1992)

# Ústav pro jazyk český – Jitka Štindlová

- **Jitka Štindlová** (1928–23. 11. 1989)
- velká zastánkyně používání počítačů v lingvistice
- 1966 uspořádala v Jílovišti mezinárodní kolokvium o mechanizaci a automatizaci v jazykovědě (velký ohlas)
- **Mechanografická laboratoř**, zal. 1967, unikátní pracoviště v humanitních vědách, využití strojů v lexikografické práci
- děrnoštitkové stroje s úpravou pro českou a slovenskou abecedu
- Štindlová zkušenosti ze zahraničí – na konci 60. let ještě spolupráce se zahraničními univerzitami
- **převod SSJČ na děrné štítky** – jeden z hlavních úkolů
- strojové zpracování Slezských písní P. Bezruč, konkordance
  - *Konkordanční a frekvenční index k Slezským písním*

# Jitka Štindlová

- **kódování češtiny** – číselný kód pro vyjádření českých znaků
- **algoritmus pro dělení slov** a zarovnání pravého okraje textu
- *Retrográdní index SSJČ*, nedočkala se vydání
- retrográdní rejstřík k *Příručnímu slovníku jazyka českého*
- 1973 vyhozena, výpočetní střediska různých podniků
- laboratoř zanikla 1975
- (*Zdeňka Sochová a Jitka Štindlová: K novým ekonomickým termínům typu „člověkohodina“, NŘ, 1953*)
  
- **Eleonora Slavíčková** – *Retrográdní morfemický slovník*, 1975, s využitím techniky Ústavu teorie informace ČSAV
  
- Jan Průcha (pedagogika), Lída Uhlířová, Jan Králík (matematik), Karel Pala (1962–1964 studijní pobyt)

# Filozofická fakulta UK

- strojový překlad, extrakce termínů
- od 1950 Katedra českého jazyka, obecného jazykozpytu a fonetiky (vedl Bohuslav Havránek)
- od 1963 Katedra obecné lingvistiky a strojového překladu (vedl Vladimír Skalička)
- **Oddělení pro teorii strojového překladu**, 1957–1962 , ved. [Petr Sgall](#)
- **Laboratoř algebraické lingvistiky**, 1968–1972, ved. **Petr Sgall**
- lingvistická skupina na MFF UK
- [Jarmila Panevová](#), [Eva Hajičová](#), [Petr Piřha](#), Zdeněk Kirschner, Bohumil Palek, Dana Konečná, Pavel Novák, Vladimír Smetáček

# Petr Sgall (\*1926)

- 1964 – *The Prague Bulletin of Mathematical Linguistics* – 2 x ročně dodnes
- 1968 **Laboratoř algebraické lingvistiky** FF UK, ale musela brzo přejít na MFF, 1972
- 1990 **Ústav formální a aplikované lingvistiky** (ÚFAL) na MFF UK
- 1990 **Ústav teoretické a počítačové lingvistiky** (ÚTKL) FF UK, 1990–92 Sgall, od 1993 Petkevič
- *Teorie informace a jazykověda*, 1964, sborník překladů
- *Cesty moderní jazykovědy*, 1964, sborník článků (Daneš, Konečná, Nebeský, Novák, Palek, Panevová, Sgall, Štindlová, Těšitelová)
- (*Sgall, P. – Zažil jsem toho dost*, 2014)

# Strojový překlad v ČSSR

- **ČSSR** – od r. 1957
- *Oddělení teorie strojového překladu FF UK* (P. Sgall, P. Novák, B. Palek)
- r. **1960** překlad angličtina – čeština na čs. počítači SAPO
  - několik desítek slov, odborné texty, ustálená slovní zásoba
- lingvisté z UK Petr Sgall, Eva Hajičová, Jarmila Panevová, Petr Piřha, Zdeněk Kirschner, *Výzkumný ústav matematických strojů*
  - projekt **APAČ**, 70. léta, čeština – angličtina
  - projekt **RUSLAN**, 80. léta, čeština – ruština

# FF/MFF UK

- **Zdeněk Kirschner** (1924–2008)
- připojil se po návratu z Tanzanie, vzápětí vyloučen z KSČ
- slovník terminologie počítačové lingvistiky, 5 jazyků (Terminologisches Wörterbuch, 1975)
- systém **MOZAIKA (MOSAIC)** – MFF UK, automatická extrakce terminologie
- strojový překlad čeština – angličtina
- **Bohumil Palek**, 1936, lingvista, sémiotik, překladatel (dva svazky díla Ch. S. Peirce)
- studoval v Leningradě statistickou lingvistiku u N. D. Andrejeva, zkušenosti přenesl do ČR
- mezivětné vztahy, anafora (*Cross-reference: a study from hyper-syntax, 1968; Základy obecné jazykovědy, 1989*)

# FF/MFF UK

- **Dana Konečná** – morfologický analyzátor, Martin Plátek, na počítači MINSK; (*Ke zkoumání češtiny z hlediska strojového překladu, 1960, SaS; První pokus se strojovým překladem v Československu, 1960, NŘ*)
- **Pavel Novák** (1932–2007) – jazykovědec, fonetik, albanista (učitelé Havránek, Skalička, Horálek), 1962–1964 byl vedoucím oddělení, 1965–1986 členem redakční rady nově založeného časopisu *Kybernetika*
- **Vladimír Smetáček**, **SEMAN** – sémantický analyzátor,
  - přirozený jazyk + sémantické rysy; tvorba tezauru (poloautomatizovaná), extrakce klíčových slov
  - korektory překlepů, dělení slov