

OSOLSOBĚ, Klára. Matematická lingvistika. In KRČMOVÁ, Marie, Radoslav VEČERKA, Jana PLESKALOVÁ a Petr KARLÍK. *Kapitoly z dějin české jazykovědné bohemistiky*. 1. vyd. Praha: Academia, 2007. s. 447-466. Mimo - humanitní vědy. ISBN 978-80-200-1523-5.

Matematická lingvistika

Ve druhé polovině XX. stol. dochází ve vědě k výrazné změně charakterizované vznikem nových „pomezních“ oborů. Tradiční odvětví vědy přijímají nové metody z oborů blízkých i relativně vzdálených. Objevují se disciplíny, které kombinují otázky lingvistické s tématy a metodami oborů, jako je matematika, psychologie, sociologie, antropologie, ale posléze i informatika nebo umělá inteligence. Vznikají tak disciplíny stojící na hranicích různých oborů, nikoli ovšem na jejich periferii.

Matematická lingvistika vykryštovala jako vědní obor na pomezí lingvistiky, matematiky, posléze i umělé inteligence (artificial intelligence – AI) a informatiky.

Matematická lingvistika se snaží o exaktní popis přirozeného jazyka opřený o matematické metody. Z určitých aspektů se dá dále rozčlenit na **komputační (počítačovou) lingvistiku** zabývající se zkoumáním a zpracováním **přirozeného jazyka** prostřednictvím počítačů a metod informatiky, **formální lingvistiku** zabývající se **formálním popisem gramatik a jazyků** a **kvantitativní (statistickou) lingvistiku** využívající statistické stochastické a pravděpodobnostní metody aplikované na přirozený jazyk (dnes rozsáhlé využití pro anotace jazykových korpusů).

V současnosti se ovšem prosazuje termín **počítačové zpracování přirozeného jazyka (natural language processing – NLP)**, který zahrnuje to, co je uvedeno pod komputační lingvistikou. Celkem samostatně se mluví o tzv. **language engineering** (doslovný překlad tohoto termínu – **jazykové inženýrství** je poněkud zatížen, neboť se takto původně v 50. letech XX. století označovala nežádoucí manipulace jazyka při různých pokusech o násilnou kodifikaci), jež pokrývá algoritmické techniky v popisu přirozeného jazyka a softwarové nástroje vznikající jejich implementací, tedy aplikace směřující ke strojovému překladu, automatickému ukládání a vyhledávání informací, tvorbě dotazovacích systémů, gramatických korektorů atp.

V souvislosti s technickým vývojem počítačů dochází od počátku 90. let XX. stol. k prudkému rozvoji korpusové lingvistiky (srv. níže a rovněž sam. kapitola Korpusová lingvistika).

Vysoká míra homonymie a synonymie výrazů a významová vágnost projevující se na všech jazykových rovinách klade explicitnímu (algoritmickému potažmo strojovému) popisu přirozeného jazyka do cesty řadu překážek. Budování nástrojů pro automatické odstraňování víceznačných interpretací jednotek přirozeného jazyka (automatická disambiguace) je tudíž jedním z prvořadých úkolů matematické lingvistiky, jehož vyřešení má dalekosáhlý význam pro rozvoj veškerých aplikací v oblasti NLP, v korpusové lingvistice, strojovém překladu, systémech směřujících k automatickému porozumění textu v libovolném neomezeném přirozeném jazyce (vyhledávání informací), v automatické analýze a syntéze řeči atd.

Ačkoliv se vznik matematické lingvistiky datuje až od druhé poloviny XX. století, můžeme říci, že některá odvětví mají své prvopočátky již v dřívějším vývoji lingvistické vědy. Počátky matematického přístupu k jazyku se v Evropě vynořují s nástupem tzv. aristotelovského racionalismu. Byl vyvolán překladem Aristotelova díla *Organon* do latiny, k němuž dal podnět Tomáš Akvinský. V jazykovědě vedlo uplatnění racionálně logických postupů ke vzniku prvních „spekulativních“ „filozofických“ gramatik, jako byla třeba *Summa grammatica* Rogera Bacona usilující postihnout „jazyk vůbec“, vlastně jazyk univerzální. Pro Baconova současníka Raimunda Lulla kolísajícího v matematizujících představách mezi kabalou a logikou byla univerzálnost jazyka paralelní s univerzálností matematické kombinatoriky, neboť na ní je podle Lulla založen jeho výrazový plán. On sám vytvářel pojmové a písmenkové tabulky a kotouče (pro kombinaci a hierarchizaci pojmů a písemných záznamů), podobně jako Giordano Bruno aj. S tímto pracovním instrumentářem se pracovalo ještě i později. Jeho aplikací na popisy, výklady a praktické využití fakt jazykových, ale i grafických, a to ani ne tak v steganografii (tj. zkoumání tajných písem), jako spíše při vytváření šifrovaných systémů, vlastně svého druhu předchůdců dnešních těsnopisných soustav. Kombinatoriky v pojmově jazykových tabulkách, klasifikačních stromech a kotoučích se užívalo až do raného novověku; nejvýraznějším reprezentantem tohoto postupu byl pravděpodobně v XVII. stol. John Wilkins a na přelomu XVII. a XVIII. stol. G. W. Leibniz. Ve filozofických koncepcích jazyka se postupně rezignovalo na vytvoření seznamu ideálních obsahů a jejich kombinací a vztahů a pozornost se obracela stále zřetelněji k logicko-matematickému kalkulu (vlastně tedy k formální syntaxi). Tento obrat lze ve filozofii sledovat od Condorceta až po Wittgensteina. Na české půdě vstoupil do tohoto myšlenkového proudu nejvýrazněji už v XVII. stol. logik, matematik a lingvista Jan Caramuel z Lobkovic, autor několika pokusů o dokonalý filozofický jazyk. Pro logickou analýzu přirozeného jazyka mělo zásadní význam dílo matematika Gottloba Fregeho.

Na možnosti používání matematických metod v lingvistice upozorňoval již v první polovině XIX. století například ruský matematik V. J. Buňakovskij a počátkem XX. století Jan Baudouin de Courtenay.

Z jazykovědců se jako první zabýval kvantitativními jevy v jazyce americký lingvista W. D. Whitney, který zkoumal frekvenci anglických hlásek. Autorem prvního frekvenčního slovníku je německý stenograf F. W. Kädig (*Häufigkeits Wörterbuch der Deutschen Sprache*, 1898). Kvantitativní lingvistiku v první polovině XX. století ovlivnili zejména ruský matematik A. A. Markov a americký lingvista německého původu George K. Zipf.

Andrej Andrejevič Markov vydal v roce 1913 statistickou analýzu textu Evžena Oněgina (Markov, 1913). Toto dílo probudilo zájem o mezioborovou spolupráci lingvistiky a matematiky. Na základě statistického zkoumání výskytu ruských hlásek a pravděpodobnosti, s jakou po sobě následují v textu, došel k závěru, že je možné předvídat pravděpodobnost jejich výskytu („markovův proces“). Šlo o první důslednou aplikaci matematické statistiky v jazykovědě, na niž pak navázaly další aplikace teorie pravděpodobnosti a teorie informace.

O kvantitativní lingvistice se někdy mluví také jako o Zipfově lingvistice podle amerického lingvisty německého původu George Kingsleyho Zipfa, z Harvardovy univerzity. Zipf zkoumal ve 20. a 30. letech XX. století relativní frekvenci hlásek, zajímal se o psychologické a fyziologické faktory ovlivňující produkci a percepci řeči.

Upozornil na vztah mezi frekvencí slov a jejich pořadím, dále na vztah mezi frekvencí slova a počtem různých slov, která tuto frekvenci mají, a na vztah mezi frekvencí slova a počtem jeho významů (Zipfovy zákony). Jeho přínos jazykovědě byl objeven s určitým zpožděním. Teprve poté, co matematikové, fyzikové a další přírodovědci odkryli ve svých oborech analogie k zákonům, které Zipf formuloval, prorazily jeho myšlenky také v lingvistice. (Podrobněji srv. např. Zipf, 1949 aj.; Uhlířová, 2002; Hřebíček, 2002.)

Z výše zmíněných tří hlavních „proudů“ (počítačová, formální, kvantitativní) shrnovaných pod střešní název matematická lingvistika má kvantitativní (statistická) lingvistika o trochu delší tradici. Na české půdě měla své předchůdce v nahodilých, ad hoc pořizovaných statistických pozorováních a závěrech, oblíbených zejména ve filologických, na excerpcei dokladů založených pracích pozitivistického zaměření. Konceptněji aplikoval tuto metodu Martin Hattala v sérii studií, v nichž se pokoušel najít jisté pravidelnosti a zákonitosti v hláskové stavbě slovanských slov, zejména v jejich konsonantické kombinatorice v násloví, a to zcela programově s pomocí matematicko-statistických postupů. Ty vedly k explikacím dosti už systematickým a svým způsobem exaktním, ale nikoli bez nebezpečí jistého mechanického přecenění kvantitativních údajů a výsledků statistického vyhodnocení zkoumaného materiálu. (Viz Srovnávací slovanská jazykověda). Zmiňujeme-li počátky kvantitativní lingvistiky postavené na pevných metodologických základech, nelze opominout práci V. Mathesia z roku 1911 sledující potencionalnost jazykových jevů. Vychází z toho, že

jazyk se neřídí „absolutními zákony“, ale že v řeči každého jednotlivce existuje kolísání „v určitých mezích a s určitou tendencí“. Jazyk jako systém funkčně diferencovaných prostředků existující ve vědomí jednotlivých mluvčích je systém pravděpodobnostní povahy (srv. Mathesius, 1911). Zajímavostí mohou být i případy interdisciplinárních úvah o využití metod matematické statistiky ve stylistice a literární teorii (Wolf, 1928). Jako samostatná lingvistická disciplína s vlastním předmětem a metodami se kvantitativní lingvistika začala u nás rozvíjet v rámci strukturalismu a její počátky jsou spojeny s pražskou školou, s pojmem kvanta a funkčního zatížení. (Vachek, 1957; Krámský, 1959; Trnka, 1937, 1951; Těšitelová, 1951 aj.).

Relativním mezníkem pro rozvoj matematické lingvistiky ve světě bylo vydání Chomského spisu o generativní teorii (*Syntactic structures* 1957), česky vyšla v polovině 60. let překladatelskou péčí Zdeňka Hlavsy (1966). Model jazyka založený na matematických metodách, jímž byla původní Chomského generativní gramatika, má dnes mnoho variant.

O zahrnutí sémantické roviny při formálním popisu přirozeného jazyka se na počátku zasloužili Fodor a Katz (Jerry Alan Fodor, Jerrold Katz, 1963), o zahrnutí roviny pragmatické se poprvé zmiňuje Yehoshua Bar-Hillel (Bar-Hillel, 1971). K vzájemnému vztahu syntaxe, sémantiky a pragmatiky srv. též Materna, Pala, Svoboda, 1976, 1979.

O aplikaci matematických metod v jazykovědě se zasloužila řada lingvistů bývalého SSSR (O. S. Achmanova, N. D. Andrejev, R. M. Frumkina, I. A. Mel'čuk, J. V. Padučeva aj.). Z východního bloku významně přispěl k rozvinutí teorie analytických modelů rumunský lingvista a matematik Solomon Marcus. Ti měli vliv na rozvoj oboru matematické lingvistiky u nás v 60. letech.

Počítačové zpracování přirozeného jazyka (Natural Language Processing, NLP) se velmi dlouho zaměřovalo především na strojový překlad – **Machine Translation** – MT (do poloviny 80. let). Od 60. let se badatelé pokouší zachytit realitu jazyka systematicky v gramaticko-logických modelech aplikovaných při strojovém zpracování přirozeného jazyka. Současnost bere gramaticko-logickou kostru jako základ, rozeznává důležitost aktuálního, zvykového a metaforického použití jazyka. Dnes se aplikace NLP orientují nejen na strojový překlad, ale zejména na tvorbu tzv. korpusových nástrojů (manažerů, automatických anotačních programů) a na programy používané v oblasti korpusové počítačové lexikografie.

K probuzení zájmu o spojení matematických metod a výzkum přirozeného jazyka, který promptně reagoval na světový trend, došlo u nás v 60. letech. Překladový sborník *Teorie informace a jazykověda* (1964) byl impulsem pro vývoj oboru a měl vliv na řadu badatelů.

Na Chomského teorii u nás poměrně velmi rychle zareagovala skupina badatelů kolem Petra Sgalla, Pavla Nováka, Dany Konečné a Bohumila Palka, kteří na FF UK v Praze koncem 50. let otevřeli Oddělení teorie strojového překladu. Jejich tehdejší studenti vyrostli v uznávané odborníky na matematickou lingvistiku (Eva Hajičová, Dana Konečná, Ladislav Nebeský, Karel Pala, Jarmila Panevová, Petr Piřha a další). V roce 1964 vychází kniha autorského kolektivu *Cesty moderní jazykovědy* s podtitulem *Jazykověda a automatizace* pod vedením Petra Sgalla. Jednotlivé oddíly se zaměřují na jazyk a techniku (strojový překlad, automatické ukládání a vyhledávání informací), algebraickou lingvistiku (formální studium a popis jazyka, generativní gramatiku, rekognoskativní gramatiku, analytické modely jazyka a modely jazykového vývoje a jazykové různosti), kvantitativní lingvistiku (kvantitativní vztahy v lexiku, význam teorie informace a kvantitativních metod pro lingvistiku), mechanizaci a automatizaci (využití počítačů) v lingvistice.

Počátkem 60. let na FF UK v Praze vzniká pojmový rámec pro formální popis přirozeného jazyka označovaný v odborné literatuře termínem funkční generativní popis – **FGP** (anglicky **FGD** – **F**unctional **G**enerative **D**escription), který je podnes rozvíjen, doplňován a obohacován. Je výsledkem spolupráce celé řady českých badatelů (Eva Benešová-Buráňová, Eva Hajičová, Květoslava Králíková, Ladislav Nebeský, Pavel Novák, Jarmila Panevová, Petr Piřha, Petr Sgall aj.), plodně čerpá z domácí lingvistické tradice, především z pojetí závislostní syntaxe, jak je nacházíme v pracích představitelů pražské lingvistické školy, ale i z pojetí Vladimíra Šmilauera, navazujícího na Tesnièreovy zásady. Kriticky vstřebává podněty, které rozvinuli ve svých pracích Vilém Mathesius, Vladimír Skalička, Miloš Dokulil, Zdeněk Hlavsa, František Daneš a další. Popis syntaktické roviny jazyka doplňuje o formálně pojatou a rozpracovanou teorii aktuálního členění věty (srv. Sgall, Hajičová, Buráňová, 1980). Tato tradice nezůstává uzavřena sama v sobě a rozvíjí se ovšem na pozadí světového lingvistického dění. Vychází z některých metod generativní syntaxe N. Chomského. Inspiruje se bádáním na poli sémantiky, jak je rozvinuli především R. Montague a B. H. Partee. Interpretace jazyka je ve **FGP** založena na integrovaném popisu syntaxe, sémantiky a pragmatiky (podrobněji srv. Sgall, 2003).

Zásadní význam v šedesátých letech měla monografie Petra Sgalla (1967) *Generativní popis jazyka a česká deklinace*. Jednotliví badatelé uveřejňovali odborné studie např. v *The Prague Bulletin of Mathematical Linguistics*, *Slově a slovesnosti*, *Kybernetice*, *Čs. informatice* a dalších (Aplikace matematiky, Metodika a technika informací).

Matematická lingvistika je původně spojena s experimenty v oblasti strojového překladu. První pokusy uskutečněné v USA r. 1954 a o rok později v SSSR byly omezeny na doslovný překlad slovo za slovo. Nicméně i ony otevřely cestu k uplatnění strojů tam, kde dosud vládl a asi bude i nadále vládnout člověk. Prvotní optimismus vystřídaloby vystřízlivění, které však nevedlo k rezignaci, ale spíše k trpělivému pokračování cestou dílčích úspěchů.

Jedním z průkopníků strojového překladu byl i americký lingvista českého původu Paul Garvin (studoval v Karlových Varech a Praze), jenž se podílel na pokusu o strojový překlad na Georgetown University ve Washingtonu. Na Harvardu pod vedením A. G. Oettingera a S. Kuna byla experimentálně ověřena tzv. prediktivní (syntaktická) analýza. Z evropských lingvistů uveďme alespoň skupinu K. Brockhause (University Münster, Kostnice, Heidelberg). První prakticky fungující systémy se objevily v Kanadě (METEO – překlady meteorologických předpovědí, projekt TAUM – Traduction automatique à l'Université de Montréal), na ně pak navázal tým na universitě v Grenoblu (skupina GETA – Groupe d'études pour la traduction automatique v čele s Bernardem Vauquoisem). Pro účely strojového překladu se v bývalém SSSR zabývali I. I. Revzin a V. J. Rozenčevjg analýzou angličtiny, skupina kolem Ju. D. Apresjana analýzou ruštiny, tým vedený O. S. Kulaginovou problémy strojového překladu z ruštiny do francouzštiny. Z dalších autorů jmenujme alespoň N. D. Andrejeva, V. V. Ivanova, I. A. Mel'čuka, D. J. Panova, S. K. Šaumjana, V. A. Uspenského. A. K. Žolkovského. Strojovému překladu se dnes věnují především velké počítačové firmy (např. IBM, SIEMENS) a samostatně firma Systran, jež vytvořila překladový systém používaný jako oficiální v EU, i univerzitní týmy v USA, Evropě, Japonsku atd. Význam pro počítačové zpracování menších evropských jazyků pro účely strojového překladu měl velký projekt Eurotran.

Počátkem 90. let byla založena Mezinárodní asociace pro strojový překlad (IAMT) se třemi samostatnými regionálními organizacemi: evropskou (European Association for Machine Translation – EAMT), americkou (Association for Machine Translation in the Americas – AMTA) a asijsko-tichomořskou (Asian-Pacific Association for Machine Translation – AAMT). Tyto organizace sdružují výzkumné ústavy, obchodní společnosti, vědecké pracovníky, odborníky z příbuzných oborů a překladatele, jejichž společným zájmem je strojový překlad. IAMT vydává svá vlastní periodika – Machine Translation, organizuje odborné semináře a jednou za dva roky pořádá mezinárodní konferenci MT Summit.

U nás se první pokus o strojový překlad z angličtiny do češtiny konal v Praze v lednu r. 1960. Skupina lingvistů Karlovy univerzity (Eva Hajičová, Zdeněk Kirschner, Jarmila Panevová, Petr Piřha, Petr Sgall) provedla ve spolupráci s Výzkumným ústavem matematických strojů (VÚMS) tento experiment na počítači SAPO české výroby. Na ně pak navázala řada projektů zaměřených na automatický překlad. V 70. letech projekt APAČ (1977-1986) jehož cílem byla automatizace překladu mezi češtinou a angličtinou, v 80. letech se projekt RUSLAN

(1987-1990) zaměřil na strojový překlad z češtiny do ruštiny. Tyto experimenty pokračovaly počátkem 90. let v rámci projektu česko-anglického strojového překladu MATRACE (1990-1992). S úspěchy a úskalími strojového překladu seznamuje čtenáře přístupnou formou kniha *Učíme stroje česky* (Sgall, Hajičová, Piřha, 1986). Více k tomuto tématu srv. Panevová, Sgall, 1980/81; Hajičová, Kirschner, Sgall, 1981.

Publikace *Cesty moderní jazykovědy*, 1964 a právě zmíněná *Učíme stroje česky* byly přínosné nejen pro rozvoj oborů, které shrnujeme pod střešní název matematická lingvistika, ale přispěly rovněž k popularizaci moderních lingvistických trendů.

Po určité přetržce způsobené nepřízní politického vývoje dochází v 60. letech k uvolnění výzkumu na poli kvantitativní lingvistiky, v jehož rámci se vzájemně doplňují a obohacují různé pohledy podmíněné generačně i koncepčně (srv. k tomu více Těšitelová, 1999). Překladový sborník *Teorie informace a jazykověda* (1964) přinesl impulsy pro bádání zaměřené na výzkum kvantitativních vztahů ve slovní zásobě, zpřístupnil novinky referující o významu teorie informace a kvantitativních metod pro lingvistiku a inspiroval teorii i praxi kvantitativní lingvistiky. Kvantitativním metodám užívaným v lingvistickém výzkumu i dalším problémům kvantitativní lingvistiky je věnována kapitola *Kvantitativní lingvistika* v knize *Cesty moderní jazykovědy* (1964).

V roce 1961 vychází první Frekvenční slovník češtiny (FSC) (Jelínek, Bečka, Těšitelová, 1961). Na půdě Ústavu pro jazyk český vzniklo dnes už neexistující **Oddělení kvantitativní lingvistiky**, u jehož zrodu stála Marie Těšitelová a které později vedl Lubomír Doležel. V průběhu 60. – 80. let rozvíjelo toto pracoviště výzkum kvantitativních charakteristik současné češtiny v její psané i mluvené podobě. 70. léta znamenají u nás určitý průlom v technických možnostech. Vzniká první počítačově čitelný korpus, z něž vzešla skupina frekvenčních slovníků (viz níže). Jeho autorkou je Marie Těšitelová (ÚJČ ČSAV).

Z prvního elektronického korpusu u nás, zpracovaného pomocí děrných štítků a čítajícího 540 000 slovních výskytů, byla čerpána data fonologická, grafématická, morfologická, slovnědruhová, lexikální a syntaktická. Ve sborníku *The Prague Studies in Mathematical Linguistics* (PSML), v odborných časopisech (*Slovo a slovesnost* aj.) a knižních sériích (např. *Glottometrika*, *Quantitative Linguistics*, u jejichž zrodu stál slovenský lingvista Gabriel Altmann) byly průběžně publikovány články řady autorů (Helena Confortiová, Lubomír Doležel, Jan Králík, Jiří Kraus, Marie Ludvíková, Iva Nebeská, Eleonora Slavičková, Jitka Štindlová, Marie Těšitelová, Ludmila Uhlířová aj.) založené na kvantitativním výzkumu

jazyka. Pod vedením Marie Těšitelové vyšla řada odborných publikací – *Otázky lexikální statistiky* (1974), *O využití statistických metod v gramatice* (1980), *Kvantitativní charakteristiky současné české publicistiky* (1982), *Kvantitativní charakteristiky současné odborné češtiny* (1983), *Kvantitativní charakteristiky současné češtiny* (1985), *O češtině v číslech* (1987), *Quantitative linguistics* (1992).

Vycházejí další specializované slovníky vytvořené pomocí počítačů a obsahující některé frekvenční charakteristiky slovní zásoby (E. Slavíčková (1975): *Retrográdní morfemický slovník češtiny*; M. Těšitelová (1980): *Frekvenční slovník současné české publicistiky*; *Frekvenční slovník současné české administrativy*, (1983): *Frekvenční slovník jazyka věcného stylu*; *Frekvenční slovník současné odborné češtiny*; J. Králík, M. Těšitelová (1986): *Retrográdní slovník současné češtiny*).

Kromě toho byla věnována pozornost praxi i teorii v oblasti výzkumu univerzálních kvantitativních vlastností přirozeného jazyka (Luděk Hřebíček, Jan Králík, Marie Königová, Ludmila Uhlířová a další). (více srv. např. Hřebíček, 2002 aj. práce uvedené níže v bibliografii). V roce 1994 byl založen časopis Mezinárodní asociace kvantitativní lingvistiky – International Quantitative Linguistics Association (IQLA) *Journal of Quantitative Linguistics*, kde zmínění odborníci také pravidelně uveřejňovali a uveřejňují výsledky své výzkumné práce.

Diferenciace akcentů v metodologických přístupech na straně jedné a nepříznivý politický vývoj na straně druhé vedl v letech 1968 - 1973 k tomu, že část badatelů (Ladislav Nebeský, Pavel Novák aj.) zůstala na FF UK, část (Petr Sgall, Eva Hajičová, Jarmila Panevová a další) založila Laboratoř algebraické lingvistiky FF UK, ale musela brzo přejít na MFF UK. Až v roce 1989 byly založeny Ústav formální a aplikované lingvistiky (ÚFAL) na MFF UK a Ústav teoretické a počítačové lingvistiky (ÚTKL) FF UK, tedy samostatná pracoviště orientovaná výhradně na studium oboru matematické lingvistiky na UK (srv. níže). Všichni badatelé přesto po celou tuto dobu s určitými omezeními pokračovali ve výzkumné i pedagogické činnosti. Plodem jejich práce byla řada statí vycházejících v domácích i zahraničních odborných periodících (PBML, Slovo a slovesnost, Čs. informatika aj.). Z monograficky zaměřených prací jmenujme alespoň kolektivní dílo Petra Sgalla a dalších (1969) *A functional approach to syntax*, (1973) *Topic, focus and generative semantics*, (1980) *Aktuální členění věty v češtině*, (1986) *The meaning of the sentence in its semantic and pragmatic aspects*, dále práce Evy Hajičové (1975) *Negace a presupozice v sémantických*

strukturách věty a Jarmily Panevové (1971) *Čas a modalita v češtině*, (1980) *Formy a funkce ve stavbě české věty*.

(Podrobné hodnocení formálního přístupu v lingvistice a místa, které v něm zaujal ve druhé polovině XX. století český výzkum srv. Hajičová, Panevová, Sgall, 1991).

Přechod Karla Paly na katedru českého jazyka, slovanské a obecné jazykovědy filozofické fakulty Univerzity Jana Evangelisty Purkyně (FF UJEP) v r. 1964 byl podnětem pro experimenty v oboru matematické lingvistiky na univerzitě v Brně. Byly zaměřeny na syntaktickou a sémantickou analýzu přirozeného jazyka a rozvíjely se v širší spolupráci s odborníky jiných oborů, například s logikem Pavlem Maternou, který u nás uváděl práce českého logika P. Tichého (srv. Tichý, 1996 a Svoboda, Jespersen, Cheyne, 2004) a informatikem Jiřím Zlatuškou. Od druhé poloviny 70. let bylo možné díky technické spolupráci s Vysokým učením technickým v Brně a později s Ústavem výpočetní techniky (ÚVT) brněnské univerzity (tehdy UJEP), ověřovat adekvátnost teoretických přístupů prvními experimenty s automatickou syntaktickou analýzou češtiny (programový systém Wander ve spolupráci s programátory Ústavu výpočetní techniky Miroslavem Benešovským, Martinem Šmídkem a Josefem Gerbrichem). Jistý průlom představují 80. léta. Na katedře českého jazyka obecné a srovnávací jazykovědy brněnské univerzity (UJEP) spolupracuje Karel Pala, Klára Osolsobě a Stanislav Franc na integrovaném morfologicko-syntaktickém analyzátoru *klara* využívajícím jazyk Prolog a aparát **DC** gramatik (Definite Clause Grammars). Články publikované v řadě A Sborníku prací filozofické fakulty univerzity J. E. Purkyně, od roku 1990 Masarykovy univerzity (SPFF UJEP, SPFF MU), referují podrobněji o jednotlivých pokusech. Monograficky jsou výsledky spolupráce odborníků zabývajících se exaktními metodami ve vztahu k přirozenému jazyku zachyceny v monografii *Logická analýza přirozeného jazyka* (Materna, Pala, Zlatuška, 1989).

Jednou z aplikací NLP, spadající do oblasti language engineering (jazykového inženýrství), je jazyková podpora v textových editorech a sázecích systémech. Patří sem jazykové korektory pravopisných překlepů (tzv. spelling-checkery), korektory pro opravu gramatických, popřípadě stylistických chyb (grammar-checkery, style-checkery), tezaury nabízející uživateli řady významově blízkých slov, programy automatického dělení slov na konci řádku, popřípadě vícejazyčné slovníky, které autor textu může během práce s editorem otvírat jako samostatná „okna“ (nejde tedy o strojový překlad, ale o elektronickou podporu pro překladatele). Práce zaměřené na formální popis české morfologie (Hajič, 1994; Osolsobě, 1996) našly uplatnění v aplikacích zaměřených na automatickou korekci českých textů.

Do oblasti aplikací NLP patří také výzkum a vývoj automatického zpracování textu (automatická indexace, automatická tvorba terminologických tezurů a automatický překlad). V této oblasti vznikla řada originálních

systemů. Většina z nich byla vyvíjena jako projektové úkoly v rámci bývalé soustavy VTEI (Informační soustava vědeckých, technických a ekonomických informací) na specializovaných pracovištích nebo v jednotlivých oborových nebo odvětvových střediscích VTEI. V 70. letech XX. století se musíme zmínit alespoň o experimentech s automatickou indexací, které prováděl J. Janoš v OBIS při závodě Turbiny podniku Škoda Plzeň (Janoš, 1976) a za zmínku stojí také jednoduchá, ale účinná metoda automatické indexace AUTIS-AI vyvinutá na konci 80. let XX. století J. Hradilem v ODIS VTEI pro uhelný průmysl v Ostravě (Hradil, 1987; Rozkopal, 1994). Od začátku 70. let byl v rámci soustavy VTEI vyvíjen i systém SEMAN – Sémantický ANalyzátor (srv. Smetáček, 1982, 1984, Uličný, 1987). Ten byl a dosud je používán k automatické popř. automatizované tvorbě tezaurových systémů. Na akademické půdě v rámci výzkumu katedry aplikované matematiky MFF UK byl vyvinut systém MOZAIKA (na Morfologickém Odvozování Založené Automatické Indexování Koherentními Agregáty) (srv. Kirschner, 1979, 1983). V souvislosti s rozpadem soustavy VTEI po roce 1989, resp. 1991, oba největší systémy (SEMAN a MOZAIKA) budované téměř 20 let fakticky zanikly. Metodologie a technologie byly však prostřednictvím autorů těchto systémů alespoň částečně přeneseny do nových projektů. Ty jsou dnes často budovány na komerční bázi. V oblasti pojmového modelování a tvorby znalostníchází byl rozvinut velmi pozoruhodný projekt v Ústavu státu a práva AVČR, který je realizován v rámci právního informačního systému LEGSYS /LexGalaxy (srv. Kořenský, Cvrček, Novák, 1999 a též <http://www.legsys.cz/>). Více informací o jednotlivých systémech i další bibliografické údaje, lze najít v přehledové studii (Schwarz, 2005 a <http://www.ikaros.cz/Clanek.asp?ID=200303002>.)

Hovoříme-li o historii oborů zahrnovaných pod střešní název matematická lingvistika, pak nemůžeme opominout skutečnost, že COLING 1982 (9. mezinárodní konference počítační lingvistiky) se konala v Praze 5.7 – 10.7. 1982 na Karlově univerzitě. Předsedkyní organizačního výboru tohoto kongresu byla E. Hajičová a jeho konání bylo umožněno tím, že je díky Jánovi Horeckému oficiálně zaštitila Slovenská akademie věd. Nešlo tu pouze o uznání zdatnosti našich odborníků, kteří na domácí půdě přednášeli a hostili své zahraniční kolegy, ale i o to, že v době, kdy účast na mezinárodních fórech byla v naší části politicky rozděleného světa podstatně omezována, umožnilo konání konference prezentovat výsledky své práce lidem, kterým by k tomu jinak nebyla dána příležitost.

Připomeňme, že už koncem 19. století byly v reakci na dosavadní zkoumání jazyka z ryze historického hlediska položeny základy strukturalistické lingvistiky. Otcem tohoto hnutí byl Ženevan Ferdinand de Saussure (1857 – 1913). Lingvistiku vracející se k Saussurovi lze charakterizovat s ohledem na její vědecko-teoretická východiska jako empirickou vědu, jejímž cílem je synchronní popis jazyka opírající se o analýzy empiricky uchopitelného materiálu jednotlivých jazyků. Z tohoto přístupu vzešla myšlenka textového korpusu. Korpusy dnes zahrnují jazykové jevy v podobě "masových dat", která lze uchovávat a zpracovávat pomocí počítačů. Tak dochází k úzkému propojení korpusově orientovaného výzkumu jazyka a počítačové lingvistiky. Z počítačové lingvistiky se začíná postupně vydělovat korpusová lingvistika, lišící se od ní orientací na masové zpracování

korpusových dat a na aplikace z něj plynoucí. Prvním moderním elektronickým korpusem byl *The Brown Corpus of Standard American English* většinou uváděný pod názvem *Brown Corpus*. Vytvořili jej W. Nelson Francis a český rodák Henry Kučera. Ačkoliv z dnešního hlediska jde o korpus velmi malý (1 milion slovních tvarů), jednalo se o první korpus v dnešním slova smyslu (elektronický, složený ze vzorků vybraných z široké škály textů tak, aby byl dodržen požadavek reprezentativnosti korpusu). V roce 1967 shrnuli Francis a Kučera výsledky analýz vycházejících z *Brown Copusu* v nyní již klasickém díle *Computational Analysis of Present-Day American English* (srv. W. Nelson Francis, Henry Kučera, 1967). Přestože byla korpusová lingvistika zpočátku vystavena značné kritice (zejména v 50. a 60. letech ze strany N. Chomského), stala se postupem času významným metodologickým proudem jazykovědy. Kromě mnoha korpusových projektů orientovaných na angličtinu se začínají budovat korpusy dalších jazyků. Velmi podrobné informace o korpusech jednotlivých jazyků a řadu dalších odkazů lze najít na <http://www.athel.com/corpus.html>. Od roku 1996 vychází International Journal of Corpus Linguistics (IJCL) přinášející širokou škálu názorů na roli korpusové lingvistiky ve výzkumu jazyka, počítačové lexikografii a NLP.

Od počátku 90. let 20. století se také u nás dynamicky rozvíjí korpusová lingvistika. Na jaře roku 1992 se sešla v Praze skupina badatelů (František Čermák, Jan Hajič, Eva Hajičová, Jan Králík, Karel Pala, Klára Osolsobě, Věra Schmiedtová, a další), kteří založili zájmové sdružení *Počítačový fond češtiny* (PFČ) (srv. podrobněji Čermák, Králík, Pala, 1992). Cílem tohoto sdružení bylo koordinovat úsilí a zajišťovat komunikaci a spolupráci odborníků, kteří mají zájem o počítačové zpracování českého jazyka. Posléze jejich snahy nabyly institucionalizované podoby. Prvním krokem byla grantová podpora (vůbec první grant nesl název „Počítačový korpus českých psaných textů“, od roku 1993 zahrnoval spolupráci odborníků univerzity Karlovy v Praze, Masarykovy univerzity v Brně a Ústavu pro jazyk český, byl úspěšně ukončen v roce 1995). Klíčový význam pak mělo v r. 1994 založení samostatného pracoviště **Ústavu Českého národního korpusu** <http://ucnk.ff.cuni.cz/> v čele s Františkem Čermákem (srv. sam. kapitola Korpusová lingvistika). Na další rozvoj bohemistiky má velký vliv budování rozsáhlých jazykových korpusů a korpusových nástrojů na straně jedné a „vytěžování“ (mining) korpusů (využívání jazykových korpusů jako zdrojů informací o jazyce) na straně druhé. Nepochybný a netrpělivě očekávaný bude význam využívání korpusů pro počítačovou lexikografii (srv. Čermák, 1999, Čermák, Klímová, Pala, Petkevič 2001). Prvním slovníkem založeným na korpusových datech v moderním slova smyslu je *Frekvenční slovník češtiny* (Čermák, Křen, 2004).

V rámci korpusové lingvistiky se dnes termínem korpus označuje rozsáhlý (vymezení rozsahu korpusu je dáno účelem, k němuž se korpus buduje) soubor počítačově čitelných (MRF – Machine Readable Form) textů složený ze souvislých textových úseků vybraných podle jistých pravidel tak, aby reprezentovaly pokud možno jazyk jako celek v celé jeho pestrosti (reprezentativnost korpusu), který obsahuje standardní reference (anotace). Tyto reference zahrnují nejrůznější metatextové informace (vzhled textu, členění na kapitoly, odstavce, typografie, ale také údaje o typu textu – informace o autorovi, dataci, žánrovém zařazení atd.) a interpretace jednotek, z nichž je text složen (jazykové značky – tagy – anotace slovnědruhovové, morfologické, syntaktické, sémantické, prozodické aj.). Iniciativu při budování standardního způsobu anotací převzala iniciativa TEI (Text Encoding Initiative). Jedná se o aktivitu sponzorovanou hlavními vědecky orientovanými asociacemi zabývajícími se využitím komputerů v humanitních vědách: ACL (*Association for Computational Linguistics*), ALLC (*the Association for Literary and Linguistic Computing*), ACH (*the Association for Computers and Humanities*). Cílem TEI je vytvoření standardní implementace pro operace s počítačově čitelnými texty. TEI za tímto účelem používá již existující formu značkovacího jazyka SGML (Standard Generalised Markup Language), popř. XML (EXtensible Markup Language). Značkovací jazyk je jakýkoli jazyk, který vkládá do textu značky vysvětlující význam nebo vzhled jednotlivých jeho částí. Vůbec první obecný značkovací jazyk byl Generalized Markup Language (GML). Na jeho základě byl vytvořen jazyk SGML, který se v minulosti stal jedním z nejrozšířenějších značkovacích jazyků. V dnešní době se nahrazuje jazykem XML – což je rozšiřitelný značkovací jazyk, jenž je zjednodušenou verzí jazyka SGML, vzniklou odstraněním chyb SGML a jeho modernizací. Vlastním příspěvkem TEI je detailní návod k použití příslušných standardů. Text jako celek se v TEI popisuje pomocí DTD (**D**ocument **T**ype **D**escription). Celá řada korpusových projektů přijala TEI za své. TEI vydává mnoho návodů pro kódování korpusových textů. EU založilo dozorčí skupinu EAGLES (*Expert Advisory Groups on Language Engineering Standards*), která má za úkol sledovat různé evropské iniciativy a pomáhat jim. Jde o to vytvořit systém anotací, v němž by na jedné straně byla brána v úvahu specifika všech evropských jazyků a na straně druhé byla zachována jednota systému. (Více k tomuto tématu srv. Blatná, Čermák, 1995 a <http://www.tei-c.org/>).

V souvislosti s korpusovou lingvistikou se do středu zájmu dostávají aplikace NLP zaměřené na automatickou anotaci (značkování) velkých korpusů. Jsou vytvářeny počítačové programy pro automatické značkování – vkládání lingvistických informací do textu – velkých jazykových korpusů. Aplikace automatické morfologické analýzy se zaměřují na automatickou lemmatizaci (přiřazení základního tvaru tzv. lemmatu textovému tvaru slova), slovnědruhovové značkování (POS – part of speech tagging) a přiřazování gramatických významů některých gramatických kategorií. Souhrnně se tento typ označuje jako gramatické

značkování (tagování, popř. anotace, ale i lemmatizace). Vzhledem k tomu, že korpusy mají být zdrojem informací pro co nejširší okruh uživatelů, se teoretické korpusové lingvistiky zamýšlejí nad tím, jak by měla lingvistická informace vkládaná do textu vypadat a čemu by měla sloužit. (Některé zásady pro vytváření anotačních schémat formuloval Geoffrey Leech, (viz Leech, 1993). Ruku v ruce s vývojem konkrétních anotačních nástrojů jdou tedy úvahy o mezích a možnostech, smyslu a účelu různých teoretických koncepcí vtělovaných do konkrétních interpretací jazykových jednotek, jimiž jsou tzv. značky (tagy) v případě slovnědruhových a gramatických značek, stromové struktury interpretující syntaktické vztahy s ohledem na sémantické, popřípadě pragmatické aspekty vyšších jednotek jazyka, značky postihující prozodické vlastnosti úseků textů, jeho sociolingvistické aj. charakteristiky atp. Klíčovou roli v tvorbě automatických korpusových analyzátorů hraje řešení problému disambiguace (zjednoznačnění homonymních jazykových jednotek na všech úrovních jazyka). Konkrétně o problémech spojených s morfologickým anotováním a následnou disambiguací českých korpusů srv. např. Petkevič, 2001.

Historie oboru matematické lingvistiky v rámci jazykovědné bohemistiky není dosud uzavřenou kapitolou. Je rovněž nemožné hodnotit z pohledu historie vědy současné dění. Nicméně se alespoň pokusíme nastínit, na kterých pracovištích se dnes dále rozvíjí to, co bylo započato v době, o níž jsme pojednali výše, přičemž není v lidských silách odpovědně zhodnotit, co přetrvá, natož pak co bude přínosem pro další rozvoj bohemistiky. (Ostatně cílem této studie není futurologie, ale pokus o jistý pohled na historii oboru matematické lingvistiky v zrcadle české bohemistiky).

V současné době bádání v oboru matematické lingvistiky rozvíjejí zejména tato akademická pracoviště:

Karlova univerzita v Praze:

Ústav formální a aplikované lingvistiky MFF UK <http://ckl.ms.mff.cuni.cz/ufal/>, v jehož rámci pracovalo v letech 2000-2004 Centrum počítačové lingvistiky MFF UK <http://ckl.mff.cuni.cz/>. Koordinovalo spolupráci pracovišť ÚFAL MFF UK, FAV ZU, ÚJČ ČAV a ve svých aktivitách a výzkumu se zaměřovalo na budování Pražského závislostního korpusu – Prague Dependency Treebank (PDT) – anotovaného na dvou syntaktických úrovních zahrnujících údaje o aktuálním členění a hlavních typech koreference (Jarmila Panevová a další), na matematické a počítačové metody (Jan Hajič a další), na rozpoznávání řeči (ve spolupráci s Josefem Psutkou a fakultou aplikovaných věd Západočeské univerzity

(FAV ZU) v Plzni) a na aplikované a experimentální systémy včetně strojového překladu (Vladislav Kuboň, Jan Cuřín, Martin Čmejrek aj.).

Teoretická východiska na straně jedné a aplikační úspěchy na straně druhé dokumentuje řada článků (např. Hajič, Hajičová, Panevová, Sgall, 1998; Hajičová, Panevová, Sgall, 2002 aj.). Řadu příspěvků publikovaných v odborném tisku je možné najít na stránce <http://ufal.mff.cuni.cz/pdt/>.

Automatickou morfologickou analýzou a disambiguací založenou na použití statistických metod se zabývá Jan Hajič (Hajič, 1994, 2002). Praktickým výsledkem jsou nejrůznější aplikace používané např. pro morfologické značkování ČNK (více o tom Hajič, Hladká, 1997), a PDT (Hajičová, Panevová, Sgall, 2002).

V PBML (<http://ufal.mff.cuni.cz/?a=pbml>) i dalších odborných časopisech a ve sbornících z domácích i mezinárodních konferencí lze najít nejnovější práce uznávaných odborníků i celé řady mladších autorů (srv. níže bibliografie, popř. www konferencí z oboru NLP).

Dalším pracovištěm Karlovy univerzity je **Ústav teoretické a počítačové lingvistiky** FF UK <http://utkl.ff.cuni.cz/>, v jehož čele stojí Vladimír Petkevič, který se na bázi širší autorské spolupráce s Karlem Olivou, Milenou Hnátkovou a dalšími věnuje problému automatické disambiguace založené na popisu pravidel fungování přirozeného jazyka. (srv. např. Oliva, Hnátková, Petkevič, 2000; Petkevič, 2001; Petkevič, 2005; Oliva, 2005).

Dalšími spolupracovníky jsou např. Alexandr Rosen (teoretický popis, strojový překlad), Hana Skoumalová (počítačová lexikografie, strojový překlad). Jejich činnost mapují články v odborném tisku a celá řada příspěvků přednesená na mezinárodních konferencích organizovaných doma i ve světě – TDS, FDSL, aj. (viz bibliografie).

Současný výzkum v oblasti strojového překladu probíhá paralelně jak na ÚFAL MFF UK, tak na Ústavu teoretické a počítačové lingvistiky FF UK. Oba ústavy spolupracují, přičemž na jedné straně se pozornost zaměřuje – vedle teoretické lingvistiky, viz zejm. Hajičová, Partee a Sgall (1998) – na strojový překlad mezi češtinou a blízkými (zejména slovanskými) jazyky (slovenštinou, polštinou, ale i litevštinou) založený na povrchové morfosyntaktické analýze (shallow MT), na druhé straně na automatický převod mezi češtinou a vzdálenějšími jazyky (především angličtina). Druhý proud sleduje dva odlišné přístupy. První je překlad založený na hloubkové syntakticko-sémantické analýze (dependency-based MT), druhý vychází z použití modelů statistického překladu (statistical MT), které se opírají o paralelní korpusy (srv. více o tomto tématu http://ckl.mff.cuni.cz/~curin/projects/mt/MT_overview.html).

Neméně významným pracovištěm je **Ústav Českého národního korpusu** FF UK, který zajišťuje od svého založení v roce 1994 budování Českého národního korpusu (ČNK), koordinaci jednotlivých korpusově orientovaných projektů a pěstování oborů korpusová lingvistika a korpusová lexikografie (srv. více sam. kap. Korpusová lingvistika a <http://ucnk.ff.cuni.cz/>). Členové pracovního kolektivu (Renata Blatná, František Čermák, Karel Kučera Věra Schmiedtová, Michal Šulc a další) publikují nejen v odborném tisku doma i ve světě, ale populárně orientovanými knižními díly i časopiseckými a novinovými články přispívají k seznámení širší veřejnosti s děním v oboru. (srv. např. Blatná, Čermák, 2005).

Na nejrůznějších grantových projektech (korpusově orientovaná lexikografie a gramatika) spolupracuje s ostatními akademickými pracovišti **Ústav pro jazyk český AVČR** <http://www.ujc.cas.cz/oddeleni/index.php> (Karel Oliva, Jan Králík, František Štícha aj.).

Problémy sémantiky ve vztahu k logice a filozofii jazyka se zabývají ve spolupráci s ÚFAL MFF UK a FI MU Jaroslav Peregrin a Pavel Materna – **Filozofický ústav AVČR** <http://www.flu.cas.cz/>.

Pozornost vztahu jazyk – kybernetika se věnuje na **Západočeské univerzitě** v Plzni <http://www.kky.zcu.cz/>, Fakultě aplikovaných věd (FAV ZU). Na počítačové rozpoznávání mluvené řeči se zaměřuje ve svých pracích Josef Pšutka (oddělení umělé inteligence), dialogové systémy a počítačové zpracování mluvené řeči zkoumá Václav Matoušek (oddělení informačních technologií).

Od poloviny 90. let XX. stol. se díky vzniku **Fakulty informatiky (FI MU)** v Brně dynamicky rozvíjí řada experimentů, které souvisejí s aplikacemi využívajícími matematických metod ke zpracování přirozeného jazyka <http://nlp.fi.muni.cz/>. V rámci **Katedry informačních technologií** rozvíjí činnost zaměřenou na počítačové zpracování přirozeného jazyka (NLP) ve dvou specializovaných laboratořích – **Centrum počítačového zpracování přirozeného jazyka** (CPZPJ) pod vedením Karla Paly a **Laboratoř dialogu a řeči** (LSD) pod vedením Ivana Kopečka .

Lingvisticky orientovaný výzkum rozvíjený na půdě CPZPJ se týká sémantické analýzy přirozeného jazyka (Aleš Horák, Aleš Hadacz), syntaktické analýzy přirozeného jazyka (Karel Pala, Pavel Smrž), evropských projektů počítačové lexikografie EuroWordNet a

BalkaNet (Karel Pala) i aplikací zaměřených na vývoj korpusových manažerů a dalších nástrojů korpusové lexikografie (Pavel Rychlý) a korpusových nástrojů pro automatickou morfologickou analýzu a z ní vycházející automatické značkování počítačových jazykových korpusů (Radek Sedláček, Marek Veber aj.). Jejich odbornou činnost mapuje celá řada příspěvků přednesených na mezinárodních konferencích doma i ve světě. Členové laboratoře spolupracují vnitrouniverzitně s Ústavem českého jazyka FF MU (budování a značkování jazykových korpusů vznikajících v rámci grantových projektů ve spolupráci s ÚČNK).

V Laboratoři dialogu a řeči (LSD) se věnuje pozornost syntéze a rozpoznávání mluvené řeči, dialogovým systémům, korpusům mluveného jazyka, počítačovým aplikacím pro handicapované (nevidomé). LSD spolupracuje vnitrofacultně s CPZPJ, dále s Katedrou informatiky a výpočetní techniky ZU v Plzni a s Ústavem radioelektroniky Fakulty elektroniky a informatiky VUT v Brně.

Nesporný význam pro rozvoj oborů zaměřených na strojové zpracování přirozeného jazyka má od roku 1998 každoročně konaná konference **Text, Speech and Dialogue (TDS)** organizovaná fakultou informatiky (FI MU) v Brně a fakultou aplikovaných věd Západočeské univerzity (FAV ZU) v Plzni a od roku 2000 také International Speech Communication Association (ISCA). Je to jediná mezinárodní konference konaná pravidelně (každoročně) v České republice (srv. níže přehled nejvýznamnějších konferencí souvisejících s oblastí strojového zpracování přirozeného jazyka a jazykového inženýrství). Je setkáním odborníků z různých zemí a oblastí, jejichž společným zájmem je právě NLP (práce s jazykovými korpusy – texty, jejich přepis do strojově čitelné podoby (machine readable form – MRF), jazyková analýza, rozpoznávání, syntéza přirozeného jazyka v jeho psané i mluvené podobě, to vše na pozadí systémů zpracování přirozeného jazyka pomocí počítačů).

Sémantickými problémy v počítačovém zpracování přirozeného jazyka se zabývá Petr Strossa, **Ústav informatiky Filozoficko-přírodovědecké fakulty Slezské univerzity** v Opavě.

Na Technické univerzitě v Liberci v rámci fakulty mechatroniky byla na **Katedře elektroniky a zpracování signálů** založena Laboratoř počítačového zpracování řeči, v níž se otázkami řečové komunikace s počítači zabývají Jan Nouza, Miroslav Halada a další.

Ačkoliv se technický rozvoj počítačů a jejich využívání ve vědě i běžném životě rozrostly do donedávna nepředstavitelných rozměrů, budou výsledky využití tohoto novodobého „nejlepšího přítele člověka“ záviset na lidských schopnostech střízlivě zvážit jeho meze a

možnosti s ohledem na úkoly, k nimž jej člověk, vědec a tedy i lingvista bude využívat. Přestože na poli formálního popisu jazyka bylo dosaženo mnoha zajímavých výsledků, zůstává i nadále otevřená otázka, do jaké míry lze systém přirozeného jazyka a proces jeho fungování explicitně a formálně popsat. Ambiciózní je pojetí, podle kterého je jádro jazykové struktury poměrně jednoduché (projektivní závislostní strom, blízký elementárním logickým strukturám, s hodnotami lexikálních a morfologických významů i syntaktických vztahů a aktuálního členění), ale její periferie je velmi rozsáhlá a zahrnuje velkou řadu kontextově podmíněných odchylek a výjimek (Sgall 2002).

Bibliografická poznámka:

Velice dobrým zdrojem pro získání rychlé informace o světově uznávaných periodických a konferencích je webová stránka digitálního archívu časopiseckých článků a příspěvků (papers) z mezinárodních konferencí oboru počítačnické lingvistiky zřízená Asociací počítačnické lingvistiky: (ACL – Association for Computational Linguistics – Anthology – A Digital Archive of Research Papers in Computational Linguistics <http://acl.ldc.upenn.edu/>). ACL je mezinárodní vědecké profesní sdružení lidí, kteří se zabývají problémy NLP. Ročně pořádá v létě konference ve významných centrech výzkumu zaměřeného na počítačnickou lingvistiku. Asociace počítačnické lingvistiky má svůj časopis – Computational Linguistics. Asociace má americkou a evropskou sekci i řadu odborů pro jednotlivé problémové okruhy. Na výše uvedené stránce lze najít Computational Linguistics – CL Journal (od r. 1980), American Journal of Computational Linguistics – ACL (od r. 1979), příspěvky z International Conference on Computational Linguistics – COLING (od r. 1965), Conference of the European Chapter of the Association for Computational Linguistics – EACL (od r. 1983), Conference on Applied Natural Language Processing – ANLP (od r. 1983), International Workshop on Natural Language Generation (od r. 1990).

Řadu informací o jednotlivých světových pracovištích, asociacích, elektronických člancích i volně dostupném softwaru se zaměřením na NLP lze najít na <http://www.ims.uni-stuttgart.de/info/FTPServer.html>.

Informace o dění v oblasti strojového překladu (bibliografie, konference, archív atd.) lze najít na stránkách The European Association for Machine Translation (EAMT) <http://www.eamt.org/>.

Cenným zdrojem pro bibliografii v oblasti computer science je webová stránka <http://www.informatik.uni-trier.de/~ley/db/>.

Účelně uspořádané informace o korpusech jednotlivých jazyků, archívech elektronických textů, korpusových nástrojích, el. dostupné literatuře i řadu bibliografických odkazů lze najít na <http://www.athel.com/corpus.html>.

Periodika a sborníky speciálně zaměřené na matematickou lingvistiku (u nás):

PBML (The Prague Bulletin of Mathematical Linguistics <http://ufal.mff.cuni.cz/?a=pbml>).

PSML (The Prague Studies in Mathematical Linguistics 1-10, 1964-1990).

Jednotlivé články k oboru je možno najít v lingvisticky orientovaných sbornících a časopisech (Slovo a slovesnost, Naše řeč aj.) a dále např. v časopisech Čs. Informatika, Kybernetika, Czechoslovak Mathematical Journal.

Některé konference s mezinárodní účastí zaměřené na matematickou lingvistiku:

International Conference on Computational Linguistics (COLING) od r. 1965.

Conference of the European Chapter of the Association for Computational Linguistics (EACL) od r. 1979.

Conference on Applied Natural Language Processing (ANLP) od r. 1983.

International Workshop on Natural Language Generation od r. 1983.

Applications of Artificial Intelligence od r. 1990.

European Conference on Formal Description of Slavic Languages od r. 1995.

European Association for Machine Translation Conference od r. 1996.

TSD (Text, Speech and Dialogue) od r. 1998.

International Conference on Language Resources and Evaluation, LREC od r. 2000 (bienále)

International Conference on Computational Linguistics and Intelligent Text Processing CICLING od r. 2000.

Conference on Communications in Computing (CIC) od r. 2000

International Conference on Artificial Intelligence and Applications (IASTED) od r. 2001

International WordNet Conference od r. 2002.

Slovníky (zejména frekvenční) zpracované za pomoci počítačů

Jelínek Jaroslav, Bečka Josef, V., Těšitelová Marie (1961): Frekvence slov, slovních druhů a tvarů v českém jazyce. Praha : SPN.

Slavičková Eleonora (1975): Retrográdní morfemický slovník češtiny s připojenými inventárními slovníky českých morfémů kořenových, prefixálních a sufixálních. Praha: Academia.

Těšitelová Marie (1980): Frekvenční slovník současné české publicistiky, Praha : Ústav pro jazyk český ČSAV .

Těšitelová Marie (1980): Frekvenční slovník současné české administrativy, Praha : Ústav pro jazyk český ČSAV .

Těšitelová Marie (1983): Frekvenční slovník jazyka věcného stylu, Praha : Ústav pro jazyk český ČSAV .

Těšitelová Marie (1983): Frekvenční slovník současné odborné češtiny, Praha : Ústav pro jazyk český ČSAV .

Králík Jan, Těšitelová Marie (1986): Retrográdní slovník současné češtiny. Praha: Academia.

Pala Karel, Všianský Jan (1994): Slovník českých synonym. Praha: Nakladatelství Lidových Novin. (2. vyd. 2001.)

Čermák František, Křen Michal (2004): Frekvenční slovník češtiny. Praha: NLN. + 1 CD-ROM.

Bibliografie:

Bar-Hillel Yehoshua (1971): Pragmatics of Natural Language. Dordrecht: Reidel.

Barnbrook Geoff (1996): Language and Computers. Edinburgh University Press.

Bémová Alena, Králíková Květoslava (1988): K otázkám automatického zpracování českého tvarosloví, Slovo a slovesnost, 49, s. 285-295.

Bémová Alena, Panevová Jarmila, Sgall Petr (1989): Znovu k významu prostých a předložkových pádů, Slovo a slovesnost, 50, s. 128-136.

Benešovský Miroslav, Sojka Petr, Šmídek Martin (1989): Programovací jazyk Wander. In: Sborník přednášek SOFSEM 1989. Ždiar, Slovensko: UVT UJEP Brno. s. 5-10.

Blatná Renata, Čermák František (eds.) (1995): Manuál lexikografie. Jinočany: H&H.

Blatná Renata, Čermák František (eds.) (2005): Jak využívat Český národní korpus. Praha: Nakladatelství Lidové noviny.

Böhmová Alena, Jarmila Panevová, Sgall Petr (1999): Syntactic Tagging: Procedure for the Transition from the Analytic to the Tectogrammatical Tree Structures. In: TSD 1999, Berlin: Springer-Verlag , s. 34-38.

Böhmová Alena, Sgall Petr (2002): The Simple Core and the Complex Periphery of Natural Language – a Formal and a Computational View. In: Proceedings of the 19th International

Conference on Computational Linguistics (COLING 2002), Morgan Kaufmann Publishers, s. 925-931.

Buráňová Eva, Hajičová Eva, Sgall Petr (2000): Tagging of very large corpora: Topic-Focus Articulation. In: Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000), Saarbrücken : Universität des Saarlandes, s. 139-144.

Confortiová Helena (1986): On the semantic analysis of prepositions from the quantitative point of view. The Prague Studies in Mathematical Linguistics, 9, s. 53-64.

Čermák František, Králík Jan, Pala Karel (1992): Počítačová lexikografie a čeština (počítačový fond češtiny). Slovo a slovesnost. 53, s. 117-124.

Čermák František (1995): Jazykový korpus: Prostředek a zdroj poznání. Slovo a slovesnost 56, s. 119-140.

Čermák František, Králík Jan, Kučera Karel (1997): Recepce současné češtiny a reprezentativnost korpusu (Výsledky a některé souvislosti jedné orientační sondy na pozadí budování Českého národního korpusu). Slovo a slovesnost, 58, s. 41-48.

Čermák František, Sgall Petr (1997): Výzkum mluvené češtiny: jeho situace a potřeby. Slovo a slovesnost, 58, s. 15-25.

Čermák František (1999): Oxfordská lexikografie přechází také plně na korpus. Slovo a slovesnost, 60, s. 136-141.

Čermák František, Klímová Jana, Petkevič Vladimír (eds.) (2000): Studie z korpusové lingvistiky. Praha: Karolinum.

Čermák František, Klímová Jana, Pala Karel, Petkevič Vladimír (2001): Design of the Czech Lexical Database. In: McEnery A., Rayson P.: Proceedings of the Corpus Linguistics Conference. University of Lancaster, s. 119–125.

Čermák František (2001): Jazyk a jazykověda, Praha: Karolinum.

Černý Jiří (1996): Dějiny lingvistiky. Olomouc: Votobia.

Čmejrek Martin, Cuřín Jan, Havelka Jiří (2002): Czech-English Dependency-based Machine Translation: Data Preparation for the Starting up Experiments, The Prague Bulletin of Mathematical Linguistics, 78, s. 103-116.

Čmejrek Martin, Cuřín Jan, Havelka Jiří (2003): Czech-English Dependency-based Machine Translation, In: Proceedings of the 10th Conference of The European Chapter of the Association for Computational Linguistics, Budapest, Hungary, s. 83-90.

Čmejrek Martin, Cuřín Jan, Havelka Jiří (2003): Treebanks in Machine Translation, In: Proceedings of The 2nd Workshop on Treebanks and Linguistic Theories, Vaxjo, Sweden, s. 209-212.

Čmejrek Martin, Cuřín Jan, Havelka Jiří, Kuboň Vladislav (2004): Building parallel bilingual syntactically annotated corpus, In: Proceedings of The 1st International Joint Conference on Natural Language Processing, Hainan Island, China, s. 141-146.

Čmejrek Martin, Cuřín Jan, Havelka Jiří (2004): Prague Czech-English Dependency Treebank: Any Hopes for a Common Annotation Scheme? In: HLT/NAACL 2004 Workshop: Frontiers in Corpus Annotation, Boston, Massachusetts, s. 47-54.

Doležel Lubomír (1963): Předběžný odhad entropie a redundance psané češtiny, Slovo a slovesnost 24, 1963, s. 165-174.

Doležel Lubomír, Průcha Jan (1964): Kombinatorické vlastnosti soustavy českých grafémů. Slovo a slovesnost, 25, s. 166-174.

Filipec Josef (1994): K dialogu o české lexikografii a lexikologii, předpočítačové i počítačové. Slovo a slovesnost, 55, s. 132-141.

Francis W. Nelson, Kučera Henry (1967): Computational Analysis of Present-Day American English. Rhode Island: Brown University Press.

Hadacz Leo, Horák, Aleš (2000) Knowledge Representation and Reasoning with Transparent Intensional Logic. In: Knowledge-Based Software Engineering, Proceedings of JCKBSE 2000, Amsterdam: IOS Press, s. 74–80.

Hajič Jan (1987): Ruslan: An MT System between Closely Related Languages, In: Proceedings of the 3rd Conference of The European Chapter of the Association for Computational Linguistics, Copenhagen, Denmark, s.113-117.

Hajič Jan, Hajičová Eva, Rosen Alexandr (1992): Machine Translation Research in Czechoslovakia. META, 37, 4, s. 802-816.

Hajič Jan, Hajičová Eva, Hnátková Milena, Kuboň Vladislav, Panevová Jarmila, Rosen Alexandr, Sgall Petr, Skoumalová Hana (1992): MATRACE – MACHINE TRANSLATION between Czech and English. In: Proceedings of the IBM Academic Initiative Projects Seminar. Praha: České vysoké učení technické, s. 75-82.

Hajič Jan (1994): Unification Morphology Grammar. Praha: Ph.D. práce MFF UK.

Hajič Jan, Hajičová Eva, Panevová Jarmila, Sgall Petr (1998): Syntax v Českém národním korpusu, Slovo a slovesnost, 59, s. 168-177.

Hajič Jan, Krbec Pavel, Květoň Pavel, Oliva Karel, Petkevič Vladimír (2001): Serial Combination of Rules and Statistics: A Case Study in Czech Tagging. In: Proceedings of the Conference of the 39th Annual Meeting of the Association for Computational Linguistics. CNRS – Institut de Recherche en Informatique de Toulouse and Université des Sciences Sociales. Toulouse, France, s. 260–267.

Hajič Jan, Hajičová Eva, Holub Martin, Pajas Petr, Sgall Petr, Vidová-Hladká Barbora, Řezníčková Veronika (2001): The current status of the Prague Dependency Treebank, In: TSD 2001. Berlin Heidelberg New York : Springer-Verlag, s. 11-20.

Hajič Jan, Pajas Petr, Vidová-Hladká Barbora, (2001): The Prague Dependency Treebank: annotation structure and support. Philadelphia: University of Pennsylvania, 2001, s. 105-114.

Hajič Jan (2001): Statistické modelování a automatická analýza přirozeného jazyka (morfologie, syntax, překlad), Bratislava: VEDA, vydavatelstvo SAV, s. 11-33.

Hajič Jan (2002): Disambiguation of Rich Inflection (Computational Morphology of Czech). Habilitation Thesis, MFF UK, Praha: Karolinum, Charles University Press.

Hajič Jan, Hajičová Eva, Panevová Jarmila, Sgall Petr (2005): Syntaktické anotování textů z Českého národního korpusu, In: Jazyky a jazykověda. Sborník k 65. narozeninám prof. Františka Čermáka, FF UK – ÚČNK, s. 213-228.

Hajičová Eva (1975): Negace a presupozice v sémantických strukturách věty. Praha: Academia.

Hajičová Eva, Sgall Petr (1975): Negation And Focus in Man – Machine Communication. In: Proceedings of the 4th International Joint Conference on Artificial Intelligence (IJCAI 1975), Tbilisi, Gruzie, SSSR, s. 443-448.

Hajičová Eva, Sgall Petr (1980): Linguistic Meaning And Knowledge Representation In: Automatic Understanding Of Natural Language. In: Proceedings of the 8th International Conference on Computational Linguistics (COLING 1980), Tokyo, Japan, s. 67-75.

Hajičová Eva, Kirschner Zdeněk, Sgall Petr (1981): Strojový překlad z angličtiny do češtiny. Slovo a slovesnost, 42, s. 61-67.

Hajičová Eva, Vrbová, J. (1981): On the Saliance of the Elements of the Stock of Shared Knowledge, Folia Linguistica, 15, s. 291-303.

Hajičová Eva, Sgall Petr (1982): Structure Of Sentence And Inferencing In Question Answering. In: Proceedings of the 1st Conference of the European Chapter of the Association for Computational Linguistics (EACL 1983), Pisa, Italy, s. 21-25.

Hajičová Eva, Panevová Jarmila (1984): Valency (Case) Frames of Verbs. In: Contributions to Functional Syntax, Semantics and Language Comprehension, Amsterdam: Benjamins – Prague : Academia, s. 147-188.

Hajičová Eva, Sgall Petr (1985): Towards An Automatic Identification Of Topic And Focus. In: Proceedings of the 2nd Conference of the European Chapter of the Association for Computational Linguistics (EACL 1985), University of Geneva, Geneva, Switzerland, s. 263-267.

- Eva Hajičová, Sgall Petr (1986): Degress of Understanding. In: Proceedings of the 11th International Conference on Computational Linguistics (COLING 1986), Bonn, Germany, Bonn : Institut für angewandte Kommunikations- und Sprachforschung, s. 184-186.
- Hajičová Eva, Oliva Karel, Sgall Petr (1987): Odkazování v gramatice a v textu, Slovo a slovesnost. 48, s. 199-212.
- Hajičová Eva, Materna Pavel, Sgall Petr (1988): Linguistic constructions in transparent intensional logic, In: Categorical Grammar, ed. by W. Buszkowski, W. Marciszewski and J. van Benthem, Amsterdam/Philadelphia : John Benjamins Publishing Co., s.283-300.
- Hajičová Eva, Sgall Petr (1989): Linguistic Aspects of AI. Artificial Intelligence in Higher Education 1989. s. 54-59.
- Hajičová Eva (1991): "Free" word order described without unnecessary complexity. Theoretical Linguistics, 17, s. 99-106.
- Hajičová Eva, Panevová Jarmila, Sgall Petr (1991): Smysl formalismu v teoretické lingvistice a úloha českého výzkumu v ní. Slovo a slovesnost, 52, s. 175–182.
- Hajičová Eva (1993): Issues of Sentence Structure and Discourse Patterns, Praha: Univerzita Karlova.
- Hajičová Eva, Sgall Petr, Skoumalová Hana (1993): Identifying Topic and Focus by an Automatic Procedure. In: Proceedings of the 6th Conference of the European Chapter of the Association for Computational Linguistics (EACL 1993), Utrecht, Netherlands, s. 178-182.
- Hajičová Eva, Rosen Alexandr (1994): Machine Readable Dictionary as a Source of Grammatical Information. In: Antonio Zampolli et al., editors, Current Issues in Computational Linguistics: In Honor of Don Walker. Kluwer Academic Publishers.
- Hajičová Eva, Hoskovec Tomáš, Sgall Petr (1995): Discourse modelling based on hierarchy of salience. The Prague Bulletin on Mathematical Linguistics 64, 1995, s. 5-24.
- Hajičová Eva, Sgall Petr, Skoumalová Hana (1995): An Automatic Procedure for Topic-Focus Identification. Computational Linguistics 21(1). s. 81-94.
- Hajičová Eva, Partee Barbara H., Sgall Petr (1998): Topic-Focus Articulation, Tripartite Structures, and Semantic Content. Dordrecht: Kluwer Academic publishers.
- Hajičová Eva, Kruijff-Korbyová Ivana, Sgall Petr (1999): Prague Dependency Treebank: Restoration of Deletions. In: TSD 1999, Berlin: Springer-Verlag , s. 44-49.
- Hajičová Eva, Panevová Jarmila, Sgall Petr (2001): Tectogramatics in corpus tagging, Amsterdam/Philadelphia : John Benjamins Publishing Company, s. 294-299.

- Hajičová Eva, Sgall Petr (2001): Topic-focus and Salience. In: Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics, Toulouse, France, s. 268-273.
- Hajičová Eva, Panevová Jarmila, Sgall Petr (2002): K nové úrovni bohemistické práce, Využití anotovaného korpusu. Část I. Slovo a slovesnost, 63, s. 161-177.
- Eva Hajičová , Jarmila Panevová, Petr Sgall (2002): K nové úrovni bohemistické práce: Využití anotovaného korpusu. Část II. Slovo a slovesnost, 63, s. 241-262.
- Hajičová Eva, Panevová Jarmila, Sgall Petr (2002): Úvod do teoretické a počítačové lingvistiky, svazek I., Teoretická lingvistika. Praha: Karolinum.
- Hladká Barbora (2000): Morphological Tagging of Czech Language. Praha: PhD. práce, MFF UK.
- Holan Tomáš, Kuboň Vladislav, Plátek Martin (2001): On Relax-ability of Word-Order by D-grammars. In: TSD 2001, Berlin : Springer Verlag, s. 159-174.
- Holan Tomáš, Kuboň Vladislav, Oliva, Karel, Plátek Martin (2001): Word-order relaxations. In: Proceedings of the 7th International Workshop on Parsing Technologies (IWPT 2001), Beijing, China, Tsinghua University Press, s. 237-240.
- Horák Aleš, Hadacz Leo (2001): Semantics of Natural Language Expressions and Inference Verification in TIL. In: Proceedings of the 3rd Workshop on Inference in Computational Semantics (ICoS-3), Siena, Italy, Università degli Studi di Siena, 2001, s. 125-130.
- Horák Aleš, Smrž Pavel (2001): Efficient Sentence Parsing with Language Specific Features: A Case Study of Czech. In: Proceedings of the the 7th International Workshop on Parsing Technologies (IWPT 2001), Beijing, China, Tsinghua University Press, s. 221-224.
- Horák Aleš, Smrž Pavel (2002): Best Analysis Selection in Inflectional Languages. In: Proceedings of the 19th International Conference on Computational Linguistics. Taipei, Taiwan : The Association for Computational Linguistics and Chinese Language Processing, s. 363-368.
- Horák Aleš, Smrž Pavel, Kadlec Vladimír (2002): Enhancing Best Analysis Selection and Parser Comparison. In: TSD 2002, Berlin: Springer-Verlag, s. 461-466.
- Horák Aleš (2002): The Normal Translation Algorithm in Transparent Intensional Logic for Czech, Brno : Masarykova Univerzita, (PhD práce).
- Horák Aleš, Pala, Karel, Smrž Pavel (2003): Lexical Semantic Networks and Ontologies in ML, Their Viewing and Authoring. In: Proceedings of the Workshop on WWW Based Communities For Knowledge Presentation, Sharing, Mining and Protection (PSMP 2003),

within CIC 2003 (International Conference on Communications in Computing). Monte Carlo Resort, Las Vegas, Nevada: CSREA Press, s. 100-105.

Horák Aleš, Smrž Pavel (2004) New Features of WordNet Editor VisDic. In: Romanian Journal of Information Science and Technology, sv. 7, č. 1-2, 2004, s. 201-214.

Horák Aleš, Smrž Pavel (2004): VisDic – Wordnet Browsing and Editing Tool. In: Global WordNet Conference 2004 Proceedings, Brno: Masarykova Univerzita, s. 136-141.

Hradil Jan (1987): Příručka pro automatickou indexaci dokumentografických informací z oblasti uhelného hornictví. Ostrava: OKD-Automatizace řízení.

Hřebíček Luděk (1996): Word associations and text. Glottometrika, 15, s. 96-101.

Fodor Jerry Alan, Katz Jerrold (1963): The structure of a semantic theory, Language, 39, s. 170-210.

Hřebíček Luděk (1992): Text in communication: supra-sentence structures. Bochum: Brockmeyer.

Hřebíček Luděk (1997): Lectures on text theory. Praha: Orientální ústav.

Hřebíček Luděk (2000): Variations in sequences (contributions to general text theory). Praha: Orientální ústav.

Hřebíček Luděk (2002): Vyprávění o lingvistických experimentech s textem. Praha: Academia.

Hřebíček Luděk (2002): Zipf's law and text, Glottometrics, 3, s. 27-38.

Chomsky Noam (orig. 1957, čes. překl. 1966): Syntaktické struktury (z angl. Syntactic structures). Praha: Academia.

Janoš Jiří (1976): Výsledky experimentu s automatickým indexováním dokumentačních záznamů odborné literatury. Praha: Státní knihovna ČSR. s. 47-104.

Karlík Petr, Nekula Marek, Pleskalová Jana (eds.) (2002): Encyklopedický slovník češtiny. Praha: Nakladatelství Lidové noviny.

Kirschner Zdeněk (1983): MOSAIC – A Method of Automatic Extraction of Significant Terms from Texts. Praha: MFF UK.

Kirschner Zdeněk (1987): APAC3-2: An English-to-Czech Machine Translation System. Explizite Beschreibung der Sprache und automatische Textverarbeitung XIV, Praha: Karlova Univerzita.

Kirschner Zdeněk (1979): K automatizovanému zpracování textu: metoda MOZAIKA. Československá informatika, 21, 3, s. 67-73.

Kirschner Zdeněk, Rosen Alexandr (1989): APAČ – An Experiment in Machine Translation. Machine Translation, 4(3), s. 177-193.

Komárek Miroslav (1964): Sur l'appréciation fonctionnelle des alternances morphologiques, Travaux linguistiques de Prague , 1, s. 145-161.

Konečná Dana (1961): Ukázka použití statistického zkoumání při přípravě strojové syntézy českého jednoduchého slovesného tvaru. Slovo a slovesnost, 22, s. 268-273.

Königová Marie (1965): K otázce statistického výběru v lingvistice. Slovo a slovesnost, 26, 161-168.

Kopeček Ivan, Smrž Pavel (1998): Navigation and Information System for Visually Impaired People. In: Proceedings of the 6th International Conference on Computers for Handicapped Persons (ICCHP 1998). Wien : Austrian Computer Society, 1998, s. 121-128.

Kopeček Ivan, Pala Karel (1998): Prosody Modelling for Syllable-Based Speech Synthesis. In: Proceedings of the IASTED International Conference on Artificial Intelligence and Soft Computing, Cancun : IASTED, 1998, s. 134-137.

Kopeček Ivan, Pala, Karel (1999): Informační systémy a porozumění přirozenému jazyku. In: Sborník mezinárodní konference RUFIS 1999. Brno : APS Brno, s. 71-76.

Kopeček Ivan, Pala Karel (2001). Syntéza řeči založená na slabičných segmentech. In: Čeština – univerzália a specifika. Brno : Masarykova Univerzita, 2001, s. 207-214.

Kopeček Ivan, Pala Karel, Straňáková-Lopatková Markéta (2001): Ambiguity Problems in Human-Computer Interaction. In: Universal Access in HCI. London: Lawrence Erlbaum Associates Publishers, sv. 3, 2001. s. 486-489.

Korvasová Květoslava, Palek Bohumil (1962): Některé vlastnosti entropie českého slovníku, Slovo a slovesnost 23, 1962, s. 58-66.

Kořenský Jan, Cvrček František, Novák František (1999): Juristická a lingvistická analýza právních textů: (právněinformatický přístup). Praha: Academia.

Králík Jan, Nebeská Iva, Těšitelová Marie (1976): On the quantitative characteristics of the Czech texts of disputed authorship – RKZ. Prague Studies in Mathematical Linguistics 5. Praha, s. 119-147.

Králík Jan, Nebeská Iva, Uhlířová Ludmila (1982): Computational Data Analysis For Syntax. In: Proceedings of the 9th International Conference on Computational Linguistics (COLING 1982), Praha : Academia, s. 391-396.

Králík Jan (1990): On some special models in quantitative linguistics, The Prague Studies in Mathematical Linguistics, 10, s. 85-94.

Králík Jan (1997): A Short Comment on Probability in Quantitative Linguistics, Journal of Quantitative Linguistics, 4,(1-3), s.126-130.

- Králík Jan (1999): Chaos versus synergie v kvantitativní lingvistice. In: Chaos, věda a filosofie. Praha : Filosofia. s. 21-28.
- Králík Jan (2003): K problému kvantifikování vágnosti z hlediska kvantitativní lingvistiky. In: Vágnost, věda a filosofie. Sborník příspěvků. (Ed. Nosek, J.). Praha : Filosofia, 2003, s. 18-24.
- Králík Jan (2005): Statistické úvahy nad rozsahem korpusu .Sborník k 65. narozeninám prof. Františka Čermáka, FF UK – ÚČNK, 2004, s. 267-271.
- Králík Jan, Šulc Michal (2005): The Representativeness of Czech corpora. *International Journal of Corpus Linguistics* 10, 3, s. 357–366.
- Králíková Květoslava (1981): Reflexivnost sloves z hlediska automatické analýzy češtiny, *Slova a slovesnost*, 42, s. 291-298.
- Krámský Josef (1959): Teorie sdělné promluvy. *Slovo a slovesnost*, 20, č. 1, s. 55-66.
- Krámský Josef (1964): K problému redundance a zvukové identity slova. *Slovo a slovesnost*, 25, s. 101-105.
- Kraus Jiří (1965): Kódování a komprese psané češtiny, *Kybernetika*, 1, s. 74-84.
- Kraus Jiří (1972): On the stylistical-semantic analysis of adjectives in journalistic style (a quantitative approach), *The Prague Studies in Mathematical Linguistics*, 4, s. 95-106.
- Kuboň Vladislav, Plátek Martin (2001): A method of accurate robust parsing for Czech, In: TSD 2001, Berlin Heidelberg New York : Springer-Verlag, s. 92-99.
- Kuboň Vladislav (2001): A method for analyzing clause complexity, *The Prague Bulletin of Mathematical Linguistics*, 38, 75, 2001, s. 5-28.
- Kuboň Vladislav (2001): Problems of robust parsing of Czech. Praha : UFAL MFF UK, (Ph.D. práce)
- Kučerová Jana (1995, 1996): The Extension of the Condition of Projectivity for Structures with Coordination, Part 1, *The Prague Bulletin of Mathematical Linguistics*, 64, s. 67-84, Part 2, *The Prague Bulletin of Mathematical Linguistics*, 65-66, s. 67-92.
- Leech Geoffrey (1993): Corpus annotation schemes, *Literary and linguistic Computing* 8 (4), s. 275-281.
- Ludvíková Marie (1972): Some quantitative aspects of the Czech syllable, *The Prague Studies in Mathematical Linguistics*, 4, s. 141-154.
- Ludvíková Marie (1978): On the occurrence of syllables in different word positions, *The Prague Studies in Mathematical Linguistics*, 6, s. 39-45.
- Machová Svatava, Panevová Jarmila, Sgall Petr (1975): A linguistic approach to information retrieval – II. In: *Information Processing and Management* 11, (5-7) s. 147-153.

- Markov Andrej Andrejevič (1913), Primer Statističeskogo isledovanija nad tekstem "Jevgenija Oněgina" ilustrujuščij svjaz ispytanij v cep, Bull. Acad. Imp. Scien. de St. Pétersbourg. Ser VI, 7,1913, s. 153-162.
- Marková Kateřina, Panevová Jarmila (2001): Opisanije dvux tipov konversivnyx par. In: Studia Philologica. Moskva : Izd.Jazyki ruskoy kul'tury, s. 202-211.
- Materna Pavel, Pala Karel, Svoboda, Aleš, (1976): An Ordered-Triple Theory of Language. Brno Studies in English 12, Brno, s.159-186.
- Materna Pavel, Pala Karel, Svoboda, Aleš, (1979): An Ordered-Triple Theory of Language Continued. Brno Studies in English 13, Brno, s. 119-165.
- Materna Pavel, Pala Karel, Svoboda Aleš (1979): Externí a interní pragmatika. In: Otázky slovanské syntaxe IV/1, Brno, s. 53-60.
- Materna Pavel, Pala Karel, Zlatuška, Jiří (1989): Logická analýza přirozeného jazyka. Praha: Academia.
- Materna, Pavel (1997): Pragmatic Meaning. In: The Logica Yearbook 1997. Praha: Filosofía, 1998. s. 65-70.
- Materna Pavel (1999) Two levels of Semantics. In: Mezi jazykem a vědomím. Praha: Filosofía, 1999, s. 33-39.
- Materna, Pavel (2000): Dva druhy propozičních postojů. Organon F, Bratislava: AVSR, 2000, s. 243-255.
- Mathesius Vilém, (1911): O potenciálnosti jevů jazykových. Praha : Král. čes. společ. nauk.
- McEney Tony, Wilson Andrew (1996): Corpus Linguistics, Edinburgh University Press.
- Mráková Eva, Sedláček, Radek (2003): From Czech Morphology through Partial Parsing to Disambiguation. In: Computational Linguistics and Intelligent Text Processing Proceedings. Berlin: Springer-Verlag, 2003. s. 126-135.
- Mráková Eva, Pala, Karel (2003): A Procedure for Semiautomatic Building Consistent Dictionary Definitions. In: Computational Linguistics and Intelligent Text Processing Proceedings. Berlin : Springer-Verlag, 2003, s. 295-303.
- Nebeská Iva (1978): On the order of clauses in a group of complex sentences expressing causal relations, The Prague Studies in Mathematical Linguistics, 6, s. 73-81.
- Nebeská Iva (1979): The quantitative characteristics of conjunctions in newspaper texts. Prague Bulletin of Mathematical Linguistics, 32, s. 57-74.
- Nebeská Iva (1982): Kvantitativní charakteristiky souvětí. In: Těšitelová M. (ed.). Kvantitativní charakteristiky současné české publicistiky. Linguistica II. Praha, s. 55-71.

- Nebeská Iva (1983): Compound/complex sentences in non-fiction texts. *Prague Studies in Mathematical Linguistics* 8, s. 53-65.
- Nebeská Iva (1983): Kvantitativní charakteristiky souvětí v psaných a mluvených odborných projevech. In: Těšitelová M. (ed.). *Psaná a mluvená odborná čeština z kvantitativního hlediska. Linguistica IV*. Praha, s. 99-120.
- Nebeská Iva (1986): A contribution to the semantics of modal verbs from the quantitative point of view, *The Prague Studies in Mathematical Linguistics*, 9, s. 53-66.
- Nebeská Iva (1990): On expressing possibility and necessity in Czech. *Prague Studies in Mathematical Linguistics*, 10, s. 75-84.
- Nebeská Iva (1991): Může být relevance postačujícím principem komunikace? *Slovo a slovesnost*, 52, s. 104-108.
- Nebeská Iva (1991): Informační model ve vývoji psycholingvistiky. *Časopis pro moderní filologii*, 73, s. 96-102.
- Nebeský Ladislav (1962): O jedné formalizaci větného rozboru. *Slovo a slovesnost*, 23, s. 104-107.
- Nebeský Ladislav (1963): O jednom algebraickém modelu jazyka. *Slovo a slovesnost*, 24, s. 231-237.
- Nebeský Ladislav (1971): Některé otázky závislostní koncepce syntaxe. *Slovo a slovesnost*, 32, s. 20-25.
- Nebeský Ladislav (1972): Užití grafů v lingvistice. *Slovo a slovesnost*, 33, s. 328-331.
- Nebeský Ladislav (1973): O pravdivostních prvcích v gramatice. *Slovo a slovesnost*, 34, s. 212-214.
- Nebeský Ladislav (1981): Matematické vlastnosti větných struktur. *Slovo a slovesnost*, 42, s. 9-15.
- Nebeský Ladislav (1982): O jazyce matematického textu. *Slovo a slovesnost*, 43, s. 88-92.
- Nebeský Ladislav (1984): Znovu o jazyce matematického textu. *Slovo a slovesnost*, 45, s. 121-127.
- Nebeský Ladislav (1987): O dvou formálních zdrojích strukturace vět. *Slovo a slovesnost*, 48, s. 190-198.
- Nebeský Ladislav (1996): On the set of all shortest paths of a given length in a connected graph. *Czechoslovak Mathematical Journal*, 46 (121), s. 155–160.
- Nebeský Ladislav, Novák Pavel (1996): Větné faktory a jejich role v analýze věty. *Slovo a slovesnost*, 57, s. 282–295.

- Nebeský Ladislav (1997): An analysis of syntactic formulae. In: Proceedings LP 96 (ed. Palek, B.). Praha: Charles University Press, s. 451–462.
- Nepil Miloslav, Popelínský Lubomír, Žáčková, Eva (2001): Part-of-Speech Tagging by Means of Shallow Parsing, ILP and Active Learning. In: Proceedings of the Third Workshop on Learning Language in Logic (LLL 2001), Strasbourg, France, s. 58-66.
- Nepil Miloslav (2001): Automated Parser Construction from a Treebank by means of TBL and ILP. In: Proceedings of the Student Research Workshop at ACL/EACL 2001, Toulouse, France, New Brunswick : Association for Computational Linguistics, s.19-24.
- Nepil Miloslav, Popelínský Lubomír (2001): Part-of-Speech Tagging by Means of ILP and Active Learning. In: Proceedings of the Workshop on Instance Selection at ECML/PKDD 2001, Freiburg, Germany, Freiburg : Albert-Ludwigs University Freiburg, 2001, s. 25-31.
- Nepil Miloslav (2001): Learning to Parse from a Treebank: Combining TBL and ILP. In: Proceedings of the 11th International Conference on Inductive Logic Programming (ILP 2001) Strasbourg, France, Berlin : Springer Verlag, s. 179-192.
- Nepil Miloslav (2003): Detecting Annotation Errors in a Corpus by Induction of Syntactic Patterns. In: TSD 2003. Berlin : Springer Verlag, s. 74-81,
- Novák Pavel, Berka Karel (1963): Výklad fonologických a gramatických pojmů s pomocí pojmů teorie množin (I. I. Revzin: Modeli jazyka. Nauka, Moskva 1962). Slovo a slovesnost, 24, s. 133–140.
- Novák Pavel, Leška Oldřich (1965): K voprosu o "strukturnom analize" jazyka. Slovo a slovesnost, 26, s. 108–112.
- Novák Pavel, Sgall Petr (1965): K voprosu o "strukturnom analize" jazyka. Slovo a slovesnost, 26, s. 315–320.
- Novák Pavel (1966): On mathematical models of linguistic objects. In: The Prague Studies in Mathematical Linguistics. 1, s. 155–157.
- Novák Pavel (1966): On the three-level approach to syntax. Travaux linguistiques de Prague , 2, s. 219–233.
- Novák Pavel (1967): K jednomu modelu stylistické složky jazykového kódování II. Slovo a slovesnost, 28, s. 36–45.
- Novák Pavel, Sgall Petr (1968): On the Prague functional approach. Travaux linguistiques de Prague , 3, s. 291–297.
- Novák Pavel (1973): O "klamu" gramatické formy. Slovo a slovesnost, 34, s. 146–149.
- Novák Pavel (1974): K lokalistickému pojetí významu pádů. Slovo a slovesnost. 35, s. 256–260.

- Novák Pavel (1977): Empirický vs. verbální postoj v lingvistice. *Slovo a slovesnost*. 38, s. 263–264.
- Novák Pavel (1978): Některé přednosti tzv. unilaterárního pojetí (jazykového) znaku. *Slovo a slovesnost*. 39, s. 215–217.
- Novák Pavel (1979): O empirických procedurách v jazykovědě. *BulRJL*, XXII, s. 47–51.
- Novák Pavel (1983): Poznámky o jazyce lingvistiky (o směřování funkčních úseků textu v lingvistice). *Slovo a slovesnost*, 44, s. 83–90.
- Novák Pavel (1991): K poválečným osudům české lingvistiky. *Slovo a slovesnost*, 52, s. 183–193.
- Novák Pavel (1995): On linear and non-linear structuring of sentences. In: Palek, B.: (ed.), *Item Order in Natural Languages. Proceedings of LP '94*. Praha : Karolinum, s. 296–301.
- Novák Pavel (1997): Relationships between homonymy and word order in a "free" word order language (Czech). In: Palek, B.: (ed.), *Typology: Prototypes, Item Orderings and Universals. Proceedings of LP '96*. Praha : Karolinum, s. 463–469.
- Novotný Miroslav (1988): *S algebrou od jazyka ke gramatice a zpět*. Praha: Academia.
- Oliva Karel (1989): A parser for Czech implemented in systems Q. In: *Explizite Beschreibung der Sprache und automatische Textverarbeitung XVI*. Praha : MFF UK.
- Oliva Karel, Petkevič Vladimír, (1998): *Phenomena Based Description of Dependency Syntax: A Survey of Ideas and Formalization*. In: *Issues of Valency and Meaning. Studies in Honour of Jarmila Panevová*, Praha : Karolinum, s. 133–153.
- Oliva Karel, Hnátková Milena, Květoň Pavel, Petkevič Vladimír (2000): *The Linguistic Basis of a Rule-Based Tagger of Czech*. In: *TSD 2000*, Berlin, Heidelberg : Springer-Verlag, s. 3–8.
- Oliva Karel (2005): Úvahy nad teoretickými základy lingvisticky adekvátní disambiguace jazykových korpusů. In: *Jazyky a jazykověda. Sborník k 65. narozeninám prof. Františka Čermáka*, FF UK – ÚČNK, 2005, s. 229-245.
- Osolsobě Klára, Pala Karel, Franc Stanislav (1987): Česká morfologie a syntax v PROLOGU. In: *Sb.semináře SOFSEM 1987*, Bratislava: VUSEIAR, s. 38-42.
- Halasová-Osolsobě Klára. (1989): *Algoritmický popis české formální morfologie substantiv a adjektiv*, SPFFMU, A 37-38, 1989-90, s.83-97.
- Osolsobě Klára, Pala Karel (1990): *Czech Stem Dictionary for IBM PC XT/AT*, Conference on Computer Lexicography, Balatonfüred, září 1990, s. 163-172.
- Osolsobě Klára (1995): *Automatické rozpoznávání a generování českých určitých číslovek a od nich odvozených číselných pojmenování na počítači*, SPFFMU A 43, 1995, s. 31-48.

- Osolsobě Klára (1996): Formale Beschreibung der tschechischen Morphologie und ihre Anwendung auf dem Gebiet der linguistischen Datenverarbeitung, SPFFBU A 44, 1996, s. 59-70.
- Osolsobě Klára (1996): Algoritmický popis české morfologie a strojový slovník češtiny. Brno: FF MU, (PhD. práce)
- Osolsobě Klára (1997): Formale beschreibung der Tschechischen morphologie. In: Uwe Junghanns, Gerhild Zybatow (Hrsg.): Formale slavistik. Frankfurt am Main : Ververt Verlag, s. 443-451.
- Osolsobě Klára, Pala Karel, Rychlý, Pavel (1998): Frekvence vzorů českých substantiv (na materiálu ČNK). SPFFBU A 46, s.77-94.
- Osolsobě Klára, Pala Karel, Rychlý, Pavel (1998): Frekvence vzorů českých sloves (na materiálu Českého národního korpusu). Slovo a slovesnost, 59, s. 265-277.
- Osolsobě Klára, Pala Karel, Sedláček Radek, Veber Marek (2002): A Procedure for Word Derivational Processes Concerning Lexicon Extension in Highly Inflected Languages, In: Proceedings of the Third International Conference on Language Resources and Evaluation, LREC, s. 1254-1259, Las Palmas de Gran Canaria : ELRA. s. 1254-1259.
- Pala Karel (1968): Náhodné generování českých vět. Slovo a slovesnost, 29, s. 45-56.
- Pala Karel (1975): Intenzionální logika a sémantika přirozeného jazyka. Slovo a slovesnost, 36, s. 205-218.
- Pala Karel (1982): O procedurální gramatice (pro češtinu). SPFFBU A 30, s. 103-122.
- Pala Karel (1984): O sémantických reprezentacích. SPFFBU A 32, s. 24-35.
- Pala Karel (1992): Počítačové zpracování češtiny, Brno: FF MU. (habil. práce rkp.)
- Pala Karel, Rychlý Pavel, Smrž Pavel (1997): DESAM – Annotated Corpus for Czech. In: Proceedings of SOFSEM 97. Heidelberg : Springer Verlag. s. 523-530.
- Pala Karel, Ševeček Pavel (1997): Valence českých sloves, SPFFBU A 45, s. 41-54.
- Pala Karel, Rychlý Pavel, Smrž Pavel (1998): Corpus Annotation in Inflectional Languages: Czech. In: 9th International Workshop on Database and Expert Systems Applications. Los Alamitos, California : IEEE Computer Society. s. 149-153.
- Pala Karel, Ševeček, Pavel (1999): Česká lexikální databáze typu WordNet (v rámci projektu EuroWordNet-2). SPFFBU A 47, s. 51-64.
- Pala Karel (2000): Počítačové zpracování přirozeného jazyka. Brno : FI MU.
- Pala Karel, Wong Shun Ha Sylvia (2002): Chinese Characters and Top Ontology in EuroWordNet. In: Proceedings of the Global WordNet Conference 2002. Mysore : Mysore University, s. 224-233.

- Pala, Karel, Sedláček Radek, Veber, Marek (2003): Relations between Inflectional and Derivation Patterns. In: Proceedings of the Workshop on Morphological Processing of Slavic Languages, Budapest: EACL, s. 1-8.
- Pala Karel, Smrž Pavel (2004): Building Czech WordNet. Romanian Journal of Information Science and Technology, sv. 7, č. 1-2, s.79-88.
- Pala Karel, Sedláček Radek (2005): Enriching WordNet with Derivational Subnets. In: CICLing 2005, s. 305-311.
- Palek Bohumil: (ed.) (1995): Proceedings of LP'94, Item Order in Natural Languages Prague August 16–18, 1994, Praha : Charles University Press.
- Palek Bohumil (1997): (ed.): Proceedings of LP'96, Typology: prototypes, item orderings and universals, Prague, August 20–22 1996, Praha : Charles University Press.
- Panevová Jarmila (1971): Ještě k teorii valence, Slovo a slovesnost, 59, s. 1-14.
- Panevová Jarmila, Benešová Eva, Sgall Petr (1971): Čas a modalita v češtině. Acta Universitatis Carolinae – Philologica, Monographia 34, Praha: Univerzita Karlova.
- Panevová Jarmila (1974-1975): On verbal frames in Functional Generative Description, The Prague Bulletin of Mathematical Linguistics 22, s. 3-40, 23, s. 17-52.
- Panevová Jarmila (1975): Rozvití předmětová a příslovečná, doplňující a určující, Naše řeč, 58, s. 61-66.
- Panevová Jarmila (1977): On verbal frames in Functional Generative Description, The Prague Bulletin of Mathematical Linguistics 28, s. 55-72.
- Panevová Jarmila (1978): Inner Participants and Free Adverbials, The Prague Studies in Mathematical Linguistics 6, s. 227-254.
- Panevová Jarmila (1979): From Tectogrammatcs to Morphemics. Transducing components of Functional Generative Description 1. Explizite Beschreibung der Sprache und automatische Textbearbeitung IV, Praha : MFF UK.
- Panevová Jarmila (1980): Formy a funkce ve stavbě české věty. Praha: Academia.
- Panevová Jarmila, Sgall Petr (1980-81): Budou stroje rozumět česky? in: ČJL 31, 1980/81, č. 4, 172-179.
- Panevová Jarmila, Sgall Petr (1987) : Machine Translation, Linguistics, And Interlingua. In: Proceedings of the 3rd Conference of the European Chapter of the Association for Computational Linguistics (EACL 1987), Copenhagen, Denmark, University of Copenhagen, s. 99-103.

- Panevová Jarmila, Skoumalová Hana (1992): Surface and Deep Cases, In: Proceedings of the 14th International Conference on Computational Linguistics (COLING 1992), Nantes, France, s. 885-889.
- Panevová Jarmila (1994): Valency Frames and the Meaning of the Sentence. In: The Prague School of Structural and Functional Linguistics (ed. by Ph. L. Luelsdorff), Linguistic and Literary Studies in Eastern Europe 41, John Benjamins, Amsterdam-Philadelphia, s. 223-243.
- Panevová Jarmila (1996): More Remarks on Control. In: Hajičová, E. aj. (eds.), Prague Linguistic Circle Papers, Vol. 2, s. 101-120.
- Panevová Jarmila, Petkevič Vladimír (1997): Agreement in Czech and Its Formal Account. In: Formale Slavistik (Uwe Junghanns, Gerhild Zybatow eds.). Leipziger Schriften zur Kultur-, Literatur-, Sprach- und Übersetzungswissenschaft, Frankfurt am Main : Vervuert Verlag, 1997, s. 321–333.
- Panevová Jarmila, Řezníčková Veronika (2001): K možnému pojetí všeobecnosti aktantu. In: ČEŠTINA – univerzália a specifika, Brno : Masarykova Univerzita, s. 139-146.
- Panevová Jarmila (2001): Některé typy chyb ve stylu odborném a žurnalistickém a možnost jejich automatického odstranění. In: TERMINA 2000, Praha : Galén, s. 40-47.
- Panevová Jarmila (2001): Problémy reflexivního zájmena v češtině, Praha: FF UK, 2001, s. 81-88.
- Vladimír Petkevič (1987): A new dependency based specification of underlying representations of sentences. In: Theoretical Linguistics, 14, s. 143-172.
- Petkevič Vladimír (1995): A new formal specification of underlying structures, Theoretical Linguistics, 21, s. 7-61.
- Petkevič Vladimír (1997): Underlying Structure of Sentence Based on Dependency. Filozofická fakulta Univerzity Karlovy : Praha.
- Petkevič Vladimír (1998): Grammatical Coreference and Agreement in Czech, The Prague Bulletin of Mathematical Linguistics, 69, s. 5–13.
- Petkevič Vladimír (1998): Special Cases of Non-Projective Constructions in the Syntax of Czech Sentence. Slavia 1998. Česká slavistika 1998. České přednášky pro XII. mezinárodní sjezd slavistů. Krakov, s. 61–66.
- Petkevič Vladimír (1999): Czech Translation of G. Orwell's '1984': Morphology and Syntactic Patterns in the Corpus. In: TSD 1999, Berlin, Heidelberg : Springer-Verlag, s. 77–82.
- Petkevič Vladimír (2001): Automatic Detection of Subject and Verbal Predicate in the Czech Translation of G. Orwell's '1984'. In: Zybatow G., U. Junghanns, G. Mehlhorn & L. Szucsich (eds.): Current Issues in Formal Slavic Linguistics. Proceedings of the Third European

Conference on Formal Description of Slavic Languages (FDSL 1999), Leipzig Frankfurt am Main : Peter Lang, s. 506–518.

Petkevič Vladimír (2001): Český národní korpus a jeho jazykové značkování [I]. In: *Čeština doma a ve světě*. Vol. 9., No.1., s. 60-73.

Petkevič Vladimír (2001): Neprojektivní konstrukce v češtině z hlediska automatické morfologické disambiguace českých textů. In: *ČEŠTINA – univerzália a specifika 2000*, Brno : Masarykova univerzita, s. 197–205.

Petkevič Vladimír (2001): Grammatical Agreement and Automatic Morphological Disambiguation of Inflectional Languages. In: *TSD 2001*, Berlin, Heidelberg : Springer-Verlag, s. 47–53.

Petkevič Vladimír (2002): Language Corpora. In: *Gjornâl Furlan des siencis (Friulian Journal of Science)* 1, 2002. Societât Sientifiche e Tecnologjiche Furlane, Udine, s. 133–151.

Petkevič Vladimír, Hnátková Milena (2002): Automatická morfologická disambiguace předložkových skupin v Českém národním korpusu. In: *ČEŠTINA – univerzália a specifika 2001*, Praha : Nakladatelství Lidové noviny, s. 243–252.

Petkevič Vladimír (2003): Subject-Predicate Agreement and Automatic Morphological Disambiguation of the Czech National Corpus. In: *Investigations into Formal Slavic Linguistics* (eds. Kosta, P., Blaszcak, J., Frasek, J., Geist, L., Žygis, M.), *Linguistik International*. Frankfurt am Main : Peter Lang; s. 315-328.

Petkevič Vladimír (2005): Za češtinu (ne)homonymní, aneb jak odstranit slovnědruhovou a morfologickou homonymii v českých korpusech. In: *Jazyky a jazykověda. Sborník k 65. narozeninám prof. Františka Čermáka, FF UK – ÚČNK, 2005*, s. 247-266.

Piřha Petr (1981): On the case frames of nouns, *The Prague Studies in Mathematical Linguistics*, 7, s. 215-224.

Piřha Petr (1987): Základní přednáška o jazyce : sborník přednášek cyklu *Kybernetické problémy přírodovědy*. Praha : SN.

Piřha Petr (1992): *Posesivní vztah v češtině*. Praha: AVED.

Plátek Martin, Sgall Petr (1978): A Scale of Context Sensitive Languages. *Applications to Natural Language Information and Control* 38, 1, s. 1-20.

Popelínský Lubomír (2003): Strojové učení a přirozený jazyk (abstrakt tutoriálu). In: *Sborník konference ZNALOSTI 2003*. Ostrava : FEI VŠB-TU Ostrava, 2003, s. 18-19.

Popelínský Lubomír (2003): Disambiguation of case suffixes in Basque. In: *Proceedings of TALN Workshop "Traitement automatique des langues minoritaires et des petites langues*. Batz-sur-Mer : ATALA, s. 213-222.

- Popelínský Lubomír (2002): Efficient relational learning from sparse data. In: Proceedings of AIMS 2002 Conference. Varna : Springer Verlag, 2002. s. 11-20.
- Popelínský Lubomír, Pavelek Tomáš (1999): Mining lemma disambiguation rules from Czech corpora. In: Proceedings of PKDD 1999, Heidelberg : Springer Verlag, s. 53-62.
- Romportl Simeon (1961): Návrh principu automatického šifrování a dešifrace gramatických příznaků českého slovesa při překládání z češtiny do češtiny, In: SbVUT, Brno, s. 160-167.
- Romportl Simeon (1970): Struktura gramatické složky slovesných tvarů určitých v češtině, Praha: Academia.
- Romportl Simeon (1997): Genealogický strom (příspěvek k metodologii evolučních jazyků), SPFFBU A 45, s. 5-17.
- Romportl Simeon (1999): Descendenční analýza, SPFFBU A 47, s. 27-32.
- Rosen Alexandr (1996): Můžete mi to přeložit na počítači? Čeština doma i ve světě, (2) s. 99-105.
- Rosen Alexandr (2001): A constraint-based approach to dependency syntax applied to some issues of Czech word order, Praha : FF UK, 2001, (PhD. práce).
- Rosen Alexandr (2002): Word order factors as constraints on feature structures. The Prague Bulletin of Mathematical Linguistics, 78, s. 5-2.
- Rozkopal Antonín (1994): Automatizovaná indexace v dokumentografických bázích dat: vybrané přístupy k řešení v ruské, české, německé a anglické jazykové oblasti se zvláštním zřetelem na metodu AIR/PHYS, Praha: UK, (diplomová práce).
- Rychlý Pavel (2000): Korpusové manažery a jejich efektivní implementace. Brno : FI MU, (PhD práce).
- Rychlý Pavel (2000): GCQP – Multiplatform Graphical User Interface to the CQP corpus manager. In: Proceedings of the 9th EURALEX International Congress, Stuttgart : Institut für Maschinelle Sprachverarbeitung, s. 149-154.
- Rychlý Pavel, Smrž Pavel, Filipenský Pavel (1999): Document Multiplicity Elimination and Corpora Management. In: Proceedings of ISAS'99. Orlando, Florida: International Institute of Informatics and Systemics, s. 231-235.
- Sedláček Radek, Smrž Pavel (2001): A New Czech Morphological Analyser AJKA. In: TSD 2001, Berlin: Springer-Verlag, s. 100-107.
- Sedláček, Radek (2004): Morphematic analyser for Czech. Brno: FI MU, (PhD. práce).
- Sgall Petr (1959): Nové otázky matematických metod v jazykovědě. Slovo a slovesnost, 20, s. 44-55.
- Sgall Petr (1963): Převodní jazyk a teorie gramatiky. Slovo a slovesnost, 24, s. 114-128.

- Sgall Petr a kol.(1964): Cesty moderní jazykovědy – Jazykověda a automatizace. Praha: Orbis.
- Sgall Petr (1964): Generativní systémy v lingvistice. Slovo a slovesnost, 25, s. 274-282.
- Sgall Petr (1964): Zur Frage der Ebenen in Sprachsystem, Travaux linguistiques de Prague, 1, s. 95-106.
- Sgall Petr (1967): Generativní popis jazyka a česká deklinace. Praha: ČSAV.
- Sgal Petr, Nebeský Ladislav, Goralčíková A., Hajičová Eva (1969): A functional approach to syntax in generative description of language. New-York: American Elsevier Publishing Company.
- Sgal Petr, Hajičová Eva (1973): Topic, focus and generative semantics. Kronberg/Taunus: Scriptor Verlag GmbH.
- Sgall Petr, Hajičová Eva, Buráňová Eva (1980): Aktuální členění v češtině. Praha: Academia.
- Sgall Petr (1982): Automatic Understanding with a Linguistically Based Knowledge Representation. ECAI 1982. s. 240-243.
- Sgall Petr (1982): Natural Language Understanding And The Perspectives Of Question Answering. In: Proceedings of the 9th International Conference on Computational Linguistics (COLING 1982), Praha : ACADEMIA, s. 357-364.
- Petr Sgall (1984): Contributions to Functional Syntax, Semantics and Language Comprehension. Praha : Academia..
- Sgall Petr ad. (1986): Úvod do syntaxe a sémantiky. Praha: Academia.
- Sgall Petr, Hajičová Eva, Piřha Petr (1986): Učíme stroje česky. Praha: Panorama.
- Sgall Petr, Hajičová Eva, Panevová Jarmila (1986): The Meaning of the Sentence in its Semantic and Pragmatic Aspects. Praha: Academia.
- Sgall Petr (1992): Underlying structure of sentences and its relations to semantics. In: Wiener Slawistischer Almanach, Sonderband 33, Festschrift für V. Ju. Rozencvejk, ed. by Tilmann Reuther, Wien: Wiener Slawistischer Almanach, Sonderband 33. s. 273-282.
- Sgall Petr, Dressler Wolfgang, Pfeiffer Oskar, Půček Michael (1995): Experimental research on Systemic Ordering. In: Theoretical Linguistics, vol. 21, s. 197-239.
- Sgall Petr, Bémová Alena, Hajičová Eva (1996): Remarks on the semantic features of cases and prepositions as related to syntax. In: Edna Andrews, Yishai Tobin, eds.: Toward a calculus of meaning, Amsterdam / Philadelphia : John Benjamins Publishing Company, s. 71-82.
- Sgall Petr (1998): Teorie valence a její formální zpracování, Slovo a slovesnost, 59, s. 15-29.

- Sgall Petr (1999): Remarks on Sentence Prosody and Topic-Focus Articulation. In: TSD 1999, Berlin: Springer-Verlag, s. 139-145.
- Sgall Petr (2002): Freedom of language: Its nature, its sources and its consequences. Prague Linguistic Circle Papers 4, Amsterdam/Philadelphia: Benjamins, s. 309-329.
- Sgall Petr (2003): Lingvistické ohlédnutí za dvacátým stoletím. In: Český jazyk a literatura, s. 157-164. SPN & Fortuna.
- Schwarz Josef (2005): Současný stav a trendy automatické indexace dokumentů – přehledová studie. Ikaros el. časopis, č. 5.
- Skoumalová Hana (1997): Czech lexicon by two-level morphology, In: Proceedings of the Second European Seminar of TELRI – "Language Applications for a Multilingual Europe", Mannheim : Kaunas, s. 123-145.
- Skoumalová Hana (1997): "A Czech morphological lexicon", In: Proceedings of the 3rd Meeting of the ACL Special Interest Group in Computational Phonology, Madrid, Spain, s. 41-47.
- Skoumalová Hana (1997): "Verb frames in the Czech hierarchical lexicon", In: TELRI Newsletter, 6, str. 18-32.
- Skoumalová Hana (1998): Derived frames and the lexicon, In: Issues of Valency and Meaning – Studies in Honour of Jarmila Panevova, Praha : Karolinum, s. 154-168.
- Skoumalová Hana (2000): "Resources for multilingual text generation in three Slavic languages". In: Proceedings of 2nd International Conference on Language Resources and Evaluation (LREC 2000), Athene, Greece, s. 1763-1768.
- Skoumalová Hana (2000): Bridge dictionaries, In: Proceedings of The 9th EURALEX International Congress, Stuttgart, BRD, s. 799-804.
- Skoumalová Hana (2001): Czech syntactic lexicon. Praha : FF UK, (PhD práce).
- Skoumalová Hana (2001): Bridge Dictionaries as Bridges Between Languages, In: International Journal of Corpus Linguistics, zvláštní číslo, 2001, s. 95-105.
- Skoumalová Hana (2002): Verb Frames Extracted from Dictionaries, The Prague Bulletin of Mathematical Linguistics, 77, s. 19-62.
- Slavíčková Eleonora (1962): Metoda morfémové analýzy založená na aplikaci teorie pravděpodobnosti. Slovo a slovesnost, 23, s. 94-104.
- Smetáček, Vladimír (1982): Sémantický analyzátor: základní pojmy a prvky (úvod do problematiky). Olomouc: Univerzita Palackého.
- Smetáček, Vladimír (1984): Sémantický analyzátor: (experimentální ověřování). Olomouc: Univerzita Palackého.

- Smrž, Pavel (1998): Hybridní systémy pro rozpoznávání vzorů. FI MU : Brno (PhD práce).
- Smrž Pavel, Horák Aleš (1998): Determining Type of TIL Construction with Verb Valency Analyser. In: Proceedings of SOFSEM 1998, Berlin : Springer-Verlag, s. 429-436.
- Smrž Pavel, Horák Aleš (2000): Large Scale Parsing of Czech. In: Proceedings of Efficiency in Large-Scale Parsing Systems Workshop, COLING 2000, Saarbrücken : Universität des Saarlandes, 2000, s. 43-50.
- Smrž Pavel (2001): Slovníková data ve formátu XML. In: Slovenčina a čeština v počítačovom spracovaní. Bratislava, Slovenská republika : VEDA – vydavateľstvo Slovenskej akadémie vied Bratislava, 2001. s. 168-180.
- Smrž Pavel (2002): Lexical Databases in XML: A Case Study of Up-Translation of the Dictionary of Literary Czech Language. In: Proceedings of the 10th EURALEX International Congress, Copenhagen, Denmark, s. 729-734.
- Smrž Pavel (2002): Storing and retrieving Wordnet database (and other structured dictionaries) in XML lexical database management system. In: Proceedings of the First International Global WordNet Conference. Mysore, India : Central Institute of Indian Languages, Department of Secondary & Higher Education, Ministry of Human Resource Development, Government of India, s 201-206.
- Smrž Pavel, Povolný Martin (2003): Lexical Databases in XML. In: Proceedings of the EACL 2003 Workshop on Language Technology and the Semantic Web: The 3rd Workshop on NLP and XML (NLPXML-2003), Budapest, Hungary, s. 49-55
- Smrž Pavel (2003): Lessons Learned from Standardization Efforts in the Context of the BalkaNet Project and the Global Wordnet Association. In: International Roadmap for Language Resources. Paris, France, s. 11-12.
- Smrž Pavel, Povolný Martin (2003): DEB – A Dictionary Editor and Browser. In: Proceedings of the 4th Papillon Workshop. Sapporo, Japan: Hokkaido University, 2003. s. 40-48.
- Smrž Pavel, Sinopalnikova Anna, Povolný Martin (2003): Thesauri and Ontologies for Digital Libraries. In: Digital Libraries: Advanced Methods and Technologies, Digital Collections. Proceedings of the 5th National Russian Research Conference (RCDL 2003). Saint-Petersburg, Russia : Saint-Petersburg State University Published Press, s. 14-17.
- Strossa Petr (1999): Vybrané kapitoly z počítačového zpracování přirozeného jazyka. Opava: Slezská univerzita.
- Svoboda Vladimír Jespersen Bjorn, Cheyne Colin (eds.) (2004): Pavel Tichý's Collected Papers in Logic and Philosophy. Praha: Filosofia.

- Ševeček, P., Morfologické programy pro češtinu: analyzátor a lemmatizátor, rkp., 1992.
- Ševeček, P., Morfologický analyzátor a lemmatizátor pro češtinu – implementace v jazyce C, Brno, 1995.
- Štícha František (1994): Čas korpusové lingvistiky. Slovo a slovesnost, 55, s. 141-145.
- Štícha František (1997): Komputační korpusy a empirická gramatika. Slovo a slovesnost, 58, s. 26-34.
- Štindlová Jitka, Mater Erich (1968): Les Machines dans la linguistique : Colloque international sur la mécanisation et l'automation des recherches linguistiques, Praha : Academia.
- Štindlová Jitka (1970): Záznam a zpracování textů pomocí strojů. Slovo a slovesnost, 31, 64-70.
- Štindlová Jitka (1971): Lexikální indexy a jejich strojové zpracování. Slovo a slovesnost, 32, 72-76.
- Teorie informace a jazykověda (1964), přel. Berka Karel, ... (et al.), Praha: Nakladatelství Československé akademie věd.
- Těšitelová Marie (1974): Otázky lexikální statistiky, Praha: Academia.
- Těšitelová Marie (1980): O využití statistických metod v gramatice, Praha: Academia.
- Těšitelová Marie (1982): Kvantitativní charakteristiky současné české publicistiky. Praha: Academia.
- Těšitelová Marie (1983): Kvantitativní charakteristiky současné odborné češtiny, Praha: Academia.
- Těšitelová Marie a kol.(1985): Kvantitativní charakteristiky současné češtiny. Praha: Academia.
- Těšitelová Marie a kol. (1987): O češtině v číslech. Praha: Academia.
- Těšitelová Marie (1992): Quantitative linguistics. Praha: Academia.
- Těšitelová Marie (1999) : K úloze časopisu Slovo a slovesnost ve vývoji české kvantitativní lingvistiky. Slovo a slovesnost, 60, s. 81–87.
- Tichý Pavel (1996): O čem mluvíme, Vybrané stati k logické sémantice, vybral a uspořádal Jaroslav Peregrin. Praha: Filosofia.
- Trnka Bohumil (1937): Pokus o vědeckou teorii a praktickou reformu těšnopisu. Praha: Filosofická fakulta University Karlovy.
- Trnka Bohumil (1948): K výstavbě fonologické statistiky. Slovo a slovesnost, 11, s. 59-64.
- Trnka Bohumil (1951): Kvantitativní lingvistika, Časopis pro moderní filologii, 34, č. 1, s. 66-74.

Trnka Bohumil (1964): General laws of phonemic combinations, in: Vachek, J.(ed), A Prague School reader in linguistics. Bloomington: Indiana University Press. s. 294-300.

Uhlířová Ludmila (1985): Kvantitativní charakteristika současné češtiny. Praha: Academia.

Ludmila Uhlířová (1996): How long are words in Czech?, Glottometrika 15, s. 134-146.

Ludmila Uhlířová (1997): Length vs. order: Word length and clause length from the perspective of word order, Journal of quantitative linguistics, 4, s. 266-275.

Uhlířová Ludmila (2002): Zipf's notion of 'economy' on the text level, Glottometrics, 3, s. 39-60.

Uličný Oldřich (1987): Automatizovaná tvorba tezauru s využitím metody SEMAN. Československá informatika, 29, č. 1, s. 16.

Vachek Josef (1940): Poznámky k fonologii českého lexika, Listy filologické, 67, s. 395-402.

Vachek Josef (1957): On the interplay of quantitative and qualitative aspects in phonemic development, Zeitschrift für Anglistik und Amerikanistik 1, s. 5-28.

Vašák Pavel (1972): Sovětská konference o matematickém zpracování textu. Slovo a slovesnost, 33, s. 250-254.

Veber Marek, Horák Aleš, Julinek Rostislav, Smrž Pavel (1999): Automatic Structuring of Written Texts. In: TSD 1999, Berlin : Springer-Verlag, s. 101-104.

Wolf František (1928): Použití počtu pravděpodobnosti k identifikaci textu. Přednáška prosloušená při slavnostní promoci v universitní aule dne 9. listopadu 1928, Brno: MU.

Žabokrtský Zdeněk, Džeroski Sašo (2001): A machine learning approach to automatic functor assignment in the Prague Dependency Treebank, The Prague Bulletin of Mathematical Linguistics, 38, 76, 2001, s. 35-44.

Zipf George Kingsly (1949): Human Behavior and the Principle of Least-Effort. Cambridge, MA: Addison-Wesley.

<http://www.legsys.cz/>

<http://www.ikaros.cz/Clanek.asp?ID=200303002>

<http://www.ff.cuni.cz/>

<http://ckl.mff.cuni.cz/>

<http://ufal.mff.cuni.cz/pdt/>

<http://utkl.ff.cuni.cz/>

<http://www.flu.cas.cz/>

<http://www.kky.zcu.cz/>

<http://nlp.fi.muni.cz/>

http://ek1.mff.cuni.cz/~curin/projects/mt/MT_overview.html

<http://ucnk.ff.cuni.cz>

<http://acl.ldc.upenn.edu/>

<http://www.ims.uni-stuttgart.de/info/FTPServer.html>

<http://www.informatik.uni-trier.de/~ley/db/>

<http://www.eamt.org/>

<http://www.athel.com/corpus.html>

<http://www.tei-c.org/>