MATHESIOVSKÉ SEMINÁRE

Euro Summer School – Vilem Mathesius Lecture Series 17

Príspevok bol publikovaný v Jazykovednom časopise (2003, ročník 54, č. 1-2, s. 129 – 133).

V dňoch 11. – 22. marca 2002 už po sedemnástykrát prebehol v Prahe prednáškový cyklus známy ako Vilem Mathesius Lecture Series, tento rok s prívlastkom Euro Summer School. Na jeho organizácii sa tradične podieľali Výskumné a školiace stredisko Viléma Mathesia pre lingvistiku a semiotiku a Ústav formálnej a aplikovanej lingvistiky Matematicko-fyzikálnej fakulty Univerzity Karlovej. Mathesiovský seminár podporili Nadácia Viléma Mathesia, Nadácia Pražského lingvistického krúžku, Európska komisia, program HESP (Higher Education Support Programme) Inštitútu pre otvorenú spoločnosť a Matematicko-fyzikálna fakulta Univerzity Karlovej.

Prednáškové cykly počas dvoch týždňov navštevovalo 41 zahraničných účastníkov zo sedemnástich krajín prevažne zo strednej a východnej Európy (Bulharsko, Chorvátsko, Maďarsko, Macedónsko, Poľsko, Rumunsko a Slovensko), z krajín bývalého Sovietskeho zväzu (Estónsko, Gruzínsko, Kirgizstan, Litva, Rusko a Ukrajina), ale aj z Nemecka, Španielska, Francúzska a Číny a nechýbali pochopiteľne ani domáci. Zo zúčastnených frekventantov prevažnú väčšinu tvorili študenti, doktorandi a profesori lingvistiky, na druhej strane aj informatici a matematici zaoberajúci sa jazykom alebo pracujúci v oblasti počítačového spracovania prirodzeného jazyka. Aj na tohtoročnom seminári vysokú úroveň prednáškových cyklov zaručovali mená známych zahraničných vedcov: Ferenc Kiefer (Maďarsko), Hans Kamp (Nemecko), Nicoletta Calzolari, Paolo Ramat (Taliansko), Robert Frank, Frederick Jelinek, Martin Kay, Barbara Partee (USA), Greville Corbett, Geoffrey Leech (Veľká Británia). Jednotlivé prednášky v angličtine mali aj českí lingvisti Eva Hajičová, Petr Sgall, Jarmila Panevová, Libuše Dušková Jarmila Panevováa informatik Jan Hajič.

Úvodné tri dni seminára boli rôznorodé tematicky i obsahovo.

Geoffrey L e e c h z Univerzity v Lancasteri sa venoval dialogickým korpusom, dialogickým systémom a anotácii rečových aktov (*Dialogue Corpora, Dialogue Systems and Speech Act Annotation*). Cyklus troch prednášok bol prezentáciou prieniku pragmatiky, korpusovej lingvistiky a počítačovej lingvistiky. Austinova teória rečových aktov sa ukázala byť aktuálnym teoretickým pozadím anotácie dialogických korpusov, ktoré majú veľmi konkrétne aplikácie v dialogických rečových systémoch. Takéto systémy sú istým druhom

komunikácie medzi človekom a strojom a poznáme ich v každodennom živote ako vyhľadávacie automaty rôznych typov informácií (o odchode vlakov, autobusov, o dopravnej situácii, počasí, resp. automaty na objednávanie cestovných lístkov a i.). Keďže dialógy tohto druhu sa označujú ako tzv. *task driven*, čiže dialógy orientované istou úlohou, ich priebeh (jednotlivé odpovede komunikantov) je do veľkej miery predvídateľný a automatizovateľný. Príkladmi automatizovaných rečových systémov sú napr. SUNDIAL, ATIS, Philips Automatic Train Timetable Information System, Verbmobil, SUN Speech Acts.

Rôzne prístupy k morfológii predstavil už tradičný prednášateľ na Mathesiovských seminároch Ferenc K i e f e r (Maďarská akadémia vied). V jeho prístupe morfológia figurovala ako autonómny jazykový podsystém s prepojením k syntaxi, fonológii a lexikónu. V rámci každej oblasti prepojenia morfológie a ďalších jazykových subsystémov potom zhrnul súčasné teórie, pričom sa sústredil na flektívnu morfológiu (lexikálno-inkrementálne teórie, lexikálno-realizačné teórie, inferenčno-inkrementálne teórie a inferenčno-realizačné teórie), derivačnú morfológiu a skladanie slov. V súvislosti s nimi rozobral aj všeobecné morfologické princípy, ako je pravidelnosť a produktívnosť, a aj typy morfologických pravidiel (morfotakticky produktívne, sémanticky a morfotakticky produktívne, sémanticky netransparentné a neproduktívne a sémanticky transparentné a neproduktívne pravidlá).

Hlavnou témou prednášok Nicolletty C a l z o l a r i (Univerzita v Pise) boli počítačové lexikóny, obzvlášť ich sémantická rovina a ich multidimenzionálny vzťah s korpusmi. V prednáškovom cykle pod názvom *Computational Lexicons and Corpus Use: between theory and practice* boli predstavené existujúce modely sémantických lexikónov, konkrétne lexikóny projektu SIMPLE, ktoré sa vypracovali pre 12 európskych jazykov a porovnali sa s lexikónmi Wordnet. Otvorili sa aj viaceré otázky interakcie lexikónov a korpusov s ohľadom na možné výskumné perspektívy (napr. sémantická anotácia korpusu), problematika viacslovných pomenovaní, čiže tzv. multiword expressions v kontexte anotácie korpusu, a predstavený bol aj súčasný stav projektu EAGLES/ISLE, ktorý sa týka viacjazyčných lexikálnych zdrojov.

Paolo R a m a t (Univerzita v Pavii) vo svojom *Úvode do lingvistickej typológie* (Introduction to linguistic typology) začal od nastolenia základnej otázky v tejto oblasti: Čo je základom typológie jazykov? Vychádzal z predpokladu, že základom porovnávania jazykov, istým „tertia comparationis“, sú noémy, čo ilustroval na príklade noémy posesívnosti v rôznych, typologicky celkom odlišných jazykoch (portugalčina, turečtina, bretónčina, jazyk Luiseño, jazyk Atabasko, ruština, nemčina, poľština, angličtina). Chcel tým potvrdiť svoju tézu, že napriek veľkej variabilnosti medzi jednotlivými jazykmi tieto jazyky nie sú

neporovnávateľné. Pristavil sa aj pri všeobecných otázkach jazykových univerzálií a morfosyntakticej typológie, v rámci ktorej predstavil syntetickosť a analytickosť ako dva póly transparentnosti vyjadrenia jazykového obsahu. Túto tézu ilustroval na príklade vývoja týchto foriem z latinčiny do rôznych románskych jazykov. Nakoniec predstavil aj istú dynamizáciu v teórii lingvistickej typológie na príklade jazykových stratégií v geneticky príbuzných jazykoch (väčšina súčasných európskych jazykov) a geneticky vzdialenejších jazykoch, tvoriacich tzv. Sprachbunde – jazykový zväz (balkánsky, mezoamerický).

Po prvých trojdielnych cykloch prednášok nasledovali jednotlivé prednášky domácich lingvistov z Ústavu formálnej a aplikovanej jazykovedy.

Príspevok Jarmily P a n e v o v e j *Valency and Functional Generative Description* bol zameraný, ako sám názov napovedá, na predstavenie všeobecných princípov teórie funkčného generatívneho opisu. Táto teória je nielen silným explikačným prístupom v rámci teoretického opisu češtiny, ale ako formalizovaný systém aj vhodným podložím automatického spracovania jazyka. Funkčný generatívny opis sa zakladá na závislostnej syntaxi, kde za stred vety sa považuje sloveso, a autorka sa aj preto zamerala na pojem valencie ako kľúčový pojem tejto teórie. Prednášateľka opísala kritériá odlíšenia medzi valenčnými pozíciami (valency slots), ako sú napr. obligatórne členy (adjuncts), a voľnými členmi (free adjuncts), ako sú napr. modifikácie. Taktiež predstavila niektoré otvorené otázky, ktoré vznikli v procese syntaktickej anotácie Pražského závislostného korpusu.

Do oblasti teórie diskurzu nás uviedla prednáška Evy Hajičovej *Topic-Focus Articulation and Discourse Patterns*. Mala tri hlavné témy:

1. informačná štruktúra vety: východisko – jadro výpovede (Topic-Focus Articulation, TFA). Pri tejto téme bola predstavená sémantická relevancia tohto teoretického prístupu, znázornené jazykové prostriedky vyjadrenia informačnej štruktúry vety v rôznych jazykoch a jej reprezentácia vo formálnom opise jazyka;
2. TFA v anotačnej schéme Pražského závislostného korpusu (PZK). Prednášateľka predstavila tri úrovne anotácie PZK: morfologickú, analytickú čiže povrchovo syntaktickú a hĺbkovo syntaktickú (tektogramatickú), ktorá zaznamenáva aj základné vlastnosti informačnej štruktúry vety, a
3. vzťahy medzi informačnou a diskurzovou štruktúrou výpovede. V tejto téme autorka uviedla pojmy: zásoba zdieľaných vedomostí (stock of shared knowledge, SSK) a hierarchia aktivácie elementov SSK.

Petr S g a l l sa vo svojej prednáške vrátil k známym pojmom pražského štrukturalizmu centra a periférie jazykového systému (*The core and periphery of language in*

linguistic description), ktoré umiestnil v kontexte pojmu príznačnosti (markedness). Hlavnými hypotézami jeho prednášky boli:

1. prototypické prostriedky vyjadrenia gramatických hodnôt sú morfémy;
2. interaktívny charakter jazyka spolu s informačnou štruktúrou vety (TFA) konštituujú aspekt syntaktickej štruktúry;
3. závislostná syntax sa zdá byť nevyhnutnou, jej elementy sú prítomné vo väčšine opisných prístupov (Lexical Functional Grammar, Head Driven Phrase Structure, X-bar teória, teória THETA rolí);
4. každá jazyková úroveň má svoju vlastnú syntax: syntax, sémantika a pragmatika nie sú tromi úrovňami jazykového systému;
5. jazyk ako systém spracovaný ľuďmi by sa mal odlišovať v komunikačnom použití od jeho kognitívnych vrstiev;
6. hĺbková syntax sa má odlišovať od morfológie, kým prvá sa zobrazuje v podobe vetných „stromov“, druhá má podobu reťazca;
7. kým centrum jazyka sa dá opísať ako systém s relatívne jednoduchým druhom základnej vzorky, veľké a komplexné periférne domény musia byť opísané pravidlami rozličného stupňa špecifickosti s nejasnými hranicami.

Prednáška Libuše D u š k o v e j *The Prague School Concept of Centre and Periphery in the Context of Similar Approaches Abroad* bola v podstate vymenovaním obsahu jednotlivých zväzkov prác Pražského lingvistického krúžku (Travaux du Cercle linguistique de Prague) s podrobnejšou charakterizáciou zväzkov, v ktorých sa vyskytuje problematika centra a periférie jazykového systému.

Prvý prednáškový týždeň uzavrel počítačový lingvista Jan H a j i č s príspevkom *From Text corpus to structural annotation*. Predstavil v ňom základné pojmy počítačovej lingvistiky (lematizácia, tokenizácia, automatická morfologická a syntaktická analýza) a podal prehľad anotačných metód na troch úrovniach anotácie korpusu: morfologickej, povrchovo syntaktickej a tektogramatickej (hĺbkovo syntaktickej). Vychádzal hlavne zo skúsenosti pri budovaní Českého národného korpusu a Pražského závislostného korpusu.

Po víkendovej prestávke prednáškový týždeň otvoril Robert F r a n k (John Hopkins University). Jeho prednáškový cyklus *The Construction of Syntactic Structure: Grammar* bol zameraný na oblasť generatívno-transformačnej syntaxe.

Prednášky Fredericka J e l i n k a, taktiež jedného zo stálych prednášateľov na Mathesiových dňoch, boli určené skôr matematicky orientovaným odborníkom v oblasti počítačovej lingvistiky. Predstavil v nich jazykové modely na automatické rozpoznávanie reči

(*Language modelling for speech recognition*). Úlohou takéhoto jazykového modelu, ktorý je súčasťou rozpoznávača reči, je určiť apriórnu pravdepodobnosť všetkých predstaviteľných výpovedí. Existujú tri typy jazykových modelov:

1. n-gramový (predvída slovo, keď pozná predchádzajúcich n-1 slov);
2. štruktúrny (predvída slovo na základe gramatickej analýzy predchádzajúcich slov) a
3. neurálny (predvída slovo pomocou umelých neurálnych sietí).

V súčasnosti je všeobecným štandardom trigramový model, ktorý predvída slovo na základe dvoch predchádzajúcich. V rámci diskusie sa otvorila aj problematika automatického odhadu štatistických parametrov, determinujúcich výkonnosť modelu z textových dát.

Častými prednášateľmi na Mathesiovských dňoch sú aj Barbara P a r t e e (Univerzita v Massachusets) a Vladimir B o r s c h e v (Univerzita v Massachusets a Ruská akadémia vied). Témou ich trojdielneho kurzu na sedemnástom Mathesiovskom seminári bola štruktúra a interpretácia posesívnych konštrukcií v rámci formálnej a lexikálnej sémantiky (*Formal and Lexical Semantics: The Interpretation of Possesives*). Hlavnými teoretickými otázkami, ktorých sa v prednáškach dotkli, bola „súhra“ lexikálnej a kompozičnej sémantiky, „diatéza“ substantív, úloha genitívnych konštrukcií ako modifikátorov a miesto „sémantických druhov“ (semantic sorts) a „ontologických axióm“ (ontological axioms) v normálnej a metonymickej interpretácii genitívnych konštrukcií.

Na prednášku Evy Hajičovej z predchádzajúceho týždňa obsahovo nadviazal Hans K a m p (Univerzita v Stuttgarte) v prednáškovom cykle *Representation, Computation and Exploitation of Information Structure*. Pohyboval sa v oblasti sémantiky, pragmatiky, teórie diskurzu a informačnej štruktúry vety.

Greville C o r b e t (Univerzita v Surrey) sa vo svojich prednáškach venoval problematike zhody z typologického hľadiska. Snažil sa načrtnúť isté spoločné východisko pre rôzne doterajšie lingvistické pohľady na charakter zhody. Uplatnil pritom pojem tzv. kánonickej zhody, vymedzenej na základe niekoľkých kritérií, ako prototypický model, od ktorého sa iné prípady zhody vo väčšej alebo menšej miere môžu odlišovať. Vo svojom prístupe vymedzil termíny: *controller* (determinujúci člen zhody), *target* (determinovaný člen zhody), *domain* (syntaktické okolie, v ktorom sa zhoda nachádza), *features* (gramatické vlastnosti v ktorých nastáva zhoda), *conditions* (faktory, ktoré ovplyvňujú zhodu, ale samy ju netvoria). Podľa charakteru týchto elementov vytvoril typológiu kánonických až exotických zhôd na príkladoch z rôznych jazykov.

Sedemnáste Mathesiovské dni po dvoch prednáškových týždňoch uzavrela príhovorom riaditeľka Centra Viléma Mathesia Eva Hajičová, ktorá sa spolu s predsedom vedeckej rady

Petrom Sgallom a koordinátorkou Libušou Brdičkovou veľkou mierou pričinila o ďalší vydarený ročník tohto podujatia. Bol nielen obsahovo rôznorodý a prednáškovo kvalitný, ale aj organizačne veľmi dobre zabezpečený. Nečudo, že Mathesiovské dni sú zaradené medzi špičkové vedecké konferencie v Európe (High Level Scientific Conferences). Každý účastník mal možnosť odniesť si z nich cennú intelektuálnu skúsenosť, písomné podklady – handouty prednášok a množstvo nových odborných kontaktov. V Čechách sa tak ustálila tradícia atraktívneho vedeckého stretnutia. Bolo by zaujímavé pouvažovať o budovaní takej tradície aj u nás.

Alexander Horák