

in južnoevropskem ženskem gibanju prve polovice 20. stoletja, v Prago pa je prinesla tudi svojo, popolnoma svežo knjigo *Frauenbewegung um 1900: über Triest nach Zagreb*. Njena knjiga pomeni za nemški govorni prostor prvo, natančno in domišljeno predstavitev pomena osebnosti Zofke Kveder, celotno znanstveno srečanje v Pragi, ki je bil prvi simpozij, v celoti posvečen

Kvedrovi, pa priložnost za novo, ponovno in drugačno branje ter vrednotenje njene književnosti in delovanja.

Vladka Tucovič

Univerza na Primorskem

Fakulteta za humanistične študije Koper
vladka.tucovic@guest.arnes.si

ESSLLI 2007 – 19. evropska poletna šola logike, jezika in informacij, Dublin, 6.–17. 8. 2007

Letošnje prizorišče poletnega srečanja študentov, raziskovalcev in strokovnjakov s področij jezikoslovja, logike in jezikovnih tehnologij je bil Trinity College v Dublinu. Poletna šola je bila izpeljana v okviru Zveze logike, jezika in informacij (FoLLI), pri izvedbi pa so sodelovali tudi Center za jezik in komunikacijske študije, evropska Zveza za računalniško jezikoslovje, Evropski raziskovalni konzorcij za informatiko in matematiko, Zveza za kognitivno znanost, Irska znanstvena fundacija in druge ustanove ter podjetja. Vodja organizacijskega odbora je bil **Carl Vogel** (Trinity College, Dublin), vodja programskega odbora pa **Tomaž Erjavec** (Ljubljana, Institut Jožef Stefan).

Tudi tokrat je bila poletna šola vsebinsko zelo bogata. V slabih štirinajstih dneh je namreč bilo mogoče izbirati med 38 predavanji, katerih glavni namen je bil seznanjati in komentirati sodobna spoznanja in probleme s področij jezikoslovja, logike in računalništva. Vsako predavanje je trajalo en teden, kar pomeni, da je obsegalo pet srečanj. Predavanja so bila glede na zahtevnos-

tno stopnjo namenjena tako tistim, ki se šele začenjajo zanimati za tovrstno raziskovanje, kot tistim, ki se že aktivno ukvarjajo z raziskovanjem in želijo svoje znanje dopolniti oz. izpopolniti. Poleg tega je bilo organiziranih tudi šest delavnic, in sicer v obliki foruma, na katerem so doktorski študentje in strokovnjaki predstavili svoje delo na določenem raziskovalnem področju in o njem razpravljali z ostalimi strokovnjaki in zainteresiranim občinstvom. Štiri delavnice so bile namenjene predstavitvi dela na področju interdisciplinarne povezave jezika in logike, po ena pa je bila s področja interdisciplinarnega sodelovanja jezikoslovja in računalništva ter logike in računalništva. V okviru poletne šole se je bilo mogoče udeležiti tudi štirih večernih predavanj, udeleženci pa so lahko svoje raziskovalno delo predstavili na študentski sekciji, ki je potekala v popoldanskem času.

Da bi bilo mogoče toliko dogodkov spraviti v dvotedensko srečanje, so bila predavanja organizirana v tri tematske sklope, in sicer *Jezik in računalništvo*, *Logika in jezik* in *Logika in računalništvo*,

ki so potekali sočasno. V sklopu *Logika in jezik* so se ukvarjali predvsem z različnimi logičnimi formulami in operacijami, ki jih je mogoče izpeljati na osnovi besed, besednih zvez in stavkov v kontekstu določenega besedila, predstavljene so bile tudi različne logične teorije, v okviru katerih se ukvarjajo s preučevanjem jezika, glavni poudarek pa je bil na prispevku logike na področji skladnje in semantike, in sicer v povezavi s korpusnim jezikoslovjem ter kognitivnimi vedami. Posebej so v tem okviru bili izpostavljeni predvsem smisel, referenca, modalni in časovni izrazi, elipse, glagoli in samostalniki ter logični algoritmi skladenjsko označenih in neoznačenih drevesnic v povezavi z logičnima pojmomoma resničnosti in neresničnosti. V sklopu *Logika in računalništvo* so bile na predavanjih predstavljene možnosti uporabe različnih logičnih teorij, spoznanj in algoritmov na področju logičnega programiranja kot vrste računalniškega programiranja, v okviru katere se problemi predvsem opisujejo v obliki navajanja dejstev in pravil. Glavni predstavnik programskih jezikov logičnega programiranja je Prolog, v katerem program predstavlja bazo znanja, ki se izvaja na podlagi postavljanja vprašanj. Posebej je bila izpostavljena povezava logike in umetne inteligence, in sicer pojma vedenje in avtomatsko sklepanje kot osnova za t. i. večagentne sisteme (*multy-agent systems*),¹ in verjetnostni modeli in algoritmi v povezavi z dialoškimi igrami.

¹ Večagentni sistemi predstavljajo alternativen način razvoja informacijskih sistemov, ki z medsebojno komunikacijo agentov omogočajo povezovanje porazdeljenih podatkov z obstoječimi sistemi, kar je uporabno predvsem na področju spletnih storitev. Tovrstni sistemi možnosti spletnih storitev razširjajo v smislu pridobivanja znanja in ontologij, posledica pa je lažja komunikacija človeka z računalnikom.

Predavanja tematskega sklopa *Jezik in računalništvo* so večinoma izpostavljala povezavo med korpusnim jezikoslovjem in semantiko, v manjši meri pa je bila pozornost namenjena tudi skladenjskemu označevanju korpusov in povezavi med različnimi jezikoslovnimi disciplinami in korpusnim jezikoslovjem. Kar se tiče povezave korpusnega jezikoslovja in semantike, je bilo največ pozornosti posvečene ontologijam (bazam leksikalnih konceptov), in sicer tako tehnološkimi problemom pri gradnji ontologij v povezavi z odvisnostjo semantičnih elementov od posameznikovih kognitivnih procesov kot različnim računalniškim aplikacijam, ki jih je mogoče ustvariti na osnovi ontologije, in njihovemu vrednotenju.

Prvi teden se je izobraževanje v sklopu *Jezik in računalništvo* začelo s predavanji **Seana Fulopa** (Kalifornijska državna univerza v Fresnu), ki je s sistemskega vidika in po posameznih jezikovnih ravninah (glasoslovje, oblikoslovje, skladnja, semantika) predstavil bistvena jezikoslovna spoznanja o značilnostih naravnih jezikov ter jih povezal z matematičnimi in logičnimi formulami in korpusnim jezikoslovjem. **Alexander Koller** (Univerza v Kolumbiji) in **Sabine Schulte im Walde** (Univerza v Stuttgartu) sta pripravila sklop predavanj, katerih glavna tema je bila t. i. korpusna semantika. Predstavila sta pomene in pomenke odnose med besedami, pri čemer sta se osredotočila na homonimijo, polisemijo, metonimijo, sinonimijo, hiper/hiponimijo, antonimijo, holonimijo in meronimijo v povezavi s korpusnim jezikoslovjem. Natančneje sta v tem okviru predstavila predvsem organizacijo besed v WordNetu in načine razdvoumljanja besednih pomenov, pri

čemer sta nekaj pozornosti namenila tudi mednarodni organizaciji Senseval, ki je nastala z namenom vrednotenja računalniških sistemov za avtomatsko razdvoumljanje besednih pomenov in vsako leto organizira delavnice, v okviru katerih se vrednoti uspešnost posameznih sistemov avtomatskega razdvoumljanja besed, in sicer tako za angleščino kot za druge jezike. Poleg tega sta se osredotočila tudi na pripisovanje semantičnih vlog v povezavi s FrameNetom in PropBankom, računalniško prepoznavanje parafraz s pomočjo učenja iz korpusov ter načine korpusnega pridobivanja podatkov o semantični podobnosti in povezanosti besed v povezavi z WordNetom in FrameNetom. **Diana Santos** (Linguatca, portugalski center za jezikovne vire) se je posvetila vrednotenju na področju procesiranja naravnih jezikov. Po opredelitvi in splošnem pregledu vrst vrednotenja (hipoteze, vir, uporabnost sistema, adekvatnost modela itd.), virov za vrednotenje ter meritev in konceptov pri vrednotenju v povezavi z veljavnostjo, zanesljivostjo in objektivnostjo se je osredotočila na posamezne načine vrednotenja, kot so MUC (*Message Understanding Conference*), TREC (delavnice za vrednotenje uspešnosti računalniških sistemov za priklic informacij), TREC QA (delavnice za vrednotenje uspešnosti računalniških sistemov, ki omogočajo pridobivanje odgovorov na vprašanja), FEMTI, BLEU (za vrednotenje uspešnosti sistemov za strojno prevajanje), in jih natančneje predstavila. **Simone Paolo Ponzetto** in **Michael Strube** (EML Research GmbH, Heidelberg) sta v sklopu predavanj predstavila Wikipedijo kot vir za aplikacije na osnovi računalniškega procesiranja naravnih jezikov, jo ovrednotila

in poudarila njeno uporabnost zlasti za ustvarjanje sklepov o moči besedilnih semantičnih povezav, pri razdvoumljanju ter računalniškem pridobivanju odgovorov na vprašanja. Posebej sta se posvetila vprašanju, kako uporabiti rudarjenje semantičnega vodenja iz Wikipedije pri računalniški eksplikaciji koreferenčnih povezav, leksikalnih verig in koherence. **Tony Veale** (University College, Dublin) je v sklopu predavanj predstavil različne leksikalne ontologije (npr. Princeton WordNet, Cyc, Sumo, Dolce, Mikrokosmos, HowNet, KorTerm idr.) kot sistem leksikalnih kategorij, ki je nastal kot kombinacija tezavra, slovarja in enciklopedije, jih ovrednotil z vidika uporabnosti in predstavil nekaj načinov, s pomočjo katerih je mogoče obstoječe ontologije obogatiti. Pri tem je izpostavil predvsem rudarjenje implicitnih pomenov (označevanje besedilnih namigov), vključevanje zunanjih virov (Wikipedija), dinamično rast leksikalnih ontologij (avtomatično posodabljanje) in kontekstualizacijo (pridobivanje novih, dodatnih kategorij na osnovi besedilnega konteksta ali korpusa) v povezavi z orodji Zeitgeist, Palimpsest in Aristotle. **Thierry Declerck** (DFKI GmbH, Saarbrücken) je svoj sklop predavanj posvetil semantičnemu označevanju multimedijskih vsebin na osnovi ontologij. Predstavil je projekt K-space, v okviru katerega se raziskovalci ukvarjajo s semantičnimi inferencami, na osnovi katerih bi bilo mogoče ustvariti računalniške sisteme za avtomatsko ali polavtomatsko označevanje in priklic multimedijskih vsebin. Podrobneje se je posvetil ontološki organizaciji značilnosti multimedijskega konteksta ter integraciji jezikovnih značilnosti, slik in videov v skupno ontološko shemo,

ki podpira sklepe kot povezave besedila in multimedijskega materiala. **Birte Lönekker-Rodman** je svoj sklop predavanj namenila figurativnemu jeziku v povezavi s strojnim prevajanjem, avtomatskim povzemanjem besedil, priklicem informacij in odgovori na vprašanja. Osredotočila se je predvsem na idiome, metonimije, metafore in humor, predstavila njihove jezikovne značilnosti v besedilno-kognitivnem kontekstu ter občinstvo seznanila s klasifikacijo, razdvoumljanjem in sklepanjem – tremi računalniškimi pristopi za avtomatsko prepoznavanje omenjenih jezikovnih razsežnosti v povezavi z WordNetom.

Drugi teden je na predavanjih tematskega sklopa *Jezik in računalništvo* **Geoffrey K. Pullum** (Univerza v Edinburgu) predstavil sodobna spoznanja na področju skladnje angleškega jezika. Osredotočil se je predvsem na *The Cambridge Grammar of the English Language* – sodobno, sinhrono in opisno slovnico angleškega jezika in jo predstavil kot teoretično osnovo za raziskovanje in praktično delo na področju računalniškega jezikoslovja, pri čemer je način obravnave skladenjskih struktur in ostale slovnične formalizme primerjal tudi z oznakami, uporabljenimi pri gradnji milijonskega korpusa Penn Treebank, ki ga sestavljajo besedila ameriškega časopisa *Wall Street Journal*. **Horacio Saggion** (Univerza v Sheffieldu) je svoj sklop predavanj namenil integraciji orodij za računalniško procesiranje naravnega jezika na področje dostopa do informacij. Predstavil je tehnološke postopke priklica informacij (*information retrieval*), luščenja informacij (*information extraction*), povzemanja besedila, pri-

dobivanja odgovorov na vprašanja in načine njihovega vrednotenja ter prosto dostopno računalniško orodje GATE za luščenje informacij in rudarjenje podatkov, ki je uporabno za različne jezike, in njegovo uporabo na področju procesiranja naravnega jezika v primerjavi z ostalimi orodji. **Ryan McDonald** (Google Inc.) in **Joakim Nivre** (Univerza v Växju, Univerza v Uppsali) sta predstavila različne metode in načine skladenjskega označevanja stavkov, in sicer tako z zgodovinskega kot sodobnega vidika. Izpostavila sta skladenjsko označevanje v obliki drevesnic, od leve proti desni in odvisnostne grafe, predstavila nekaj algoritmov za skladenjsko označevanje stavkov, rezultate testa uspešnosti posameznih orodij za skladenjsko označevanje korpusov, ki je bil narejen za vse jezike, ki že imajo oblikovano tovrstno orodje, vključno s slovensko odvisnostno drevesnico, ter med trenutno dostopnimi orodji za skladenjsko označevanje izpostavila zlasti orodji MSTparser in MaltParser. **Paul Piwek** (Open University) in **Rodger Kibble** (Univerza v Londonu, Goldsmiths College) sta se na predavanjih posvetila dialoškim igram. Predstavila sta jih z zgodovinskega in metodološkega vidika, pri čemer sta posebej opozorila na njihovo odvisnost od širše javnosti in na strukture, ki nastanejo, ko v dialoški igri sodelujeta dva ali več udeležencev. Pri tem sta se opirala na spoznanja konverzacijske analize ter analize z uporabo spoznanj propozicijske in predikatne logike. **Michael Carl** (Inštitut za uporabno raziskovanje informacij, Saarbrücken) je svoj sklop predavanj namenil zgodovinskemu razvoju metod strojnega prevajanja, in sicer od začetkov 70. let 20. stoletja do danes, s poudarkom na t. i.

klasičnih pristopih k strojnemu prevajanju in sodobnih pristopih, pri katerih se medsebojno povezujejo in dopolnjujejo metode na osnovi pravil in metode na osnovi analize podatkov. Klasifikacijske sheme za sisteme strojnega prevajanja je predstavil na primeru sistema strojnega prevajanja METIS-II. **Paul Buitelaar** (DFKI GmbH, Saarbrücken) in **Philipp Cimiano** (Institut AIFB, Karlsruhe) sta se na predavanjih posvetila interakciji med ontologijami, ki izvirajo iz reprezentacij védenja in so nastale v kontekstu t. i. semantične mreže (*semantic web*), in ontologijami, ki so nastale na teoretični podlagi leksikalne semantike (npr. WordNet, FrameNet), ter njihovi vlogi pri razumevanju naravnega jezika. Izpostavila sta predvsem vlogo, ki jo imajo ontologije pri razdvoumljanju, diskurzni in časovni interpretaciji podatkov ter luščenju informacij (*information extraction*) in njihovi integraciji, pri odgovorih na vprašanja, dialoških sistemih in strojnem prevajanju. **Mark Stevenson** (Univerza v Sheffieldu) in **Roman Yangarber** (Univerza v Helsinkih) sta predavala o luščenju informacij (*information extraction*) in uporabnosti tovrstnega jezikovnotehnološkega postopka v povezavi z rudarjenjem podatkov v besedilih in semantično mrežo. Osredotočila sta se predvsem na razvoj sistemov za luščenje informacij s pomočjo strojnega učenja (*machine learning*) in predstavila t. i. šibko nadzorovane algoritme (*weakly supervised algorithms*) kot pomoč pri računalniškem luščenju informacij o entiteti, koreferenci in odnosih med entitetami. **Martin Emms** in **Saturnino Luz** sta predavanja posvetila tehnikam strojnega učenja (*machine learning*). Predstavila sta osnovne koncepte oz. načine učenja,

pri čemer sta nadzorovane načine strojnega učenja predstavila na primerih klasifikacije določenega besedila in razdvoumljanja besednih pomenov, nenadzorovane načine strojnega učenja pa s pomočjo metode t. i. vzorčenja (*clustering*). Nekaj pozornosti sta namenila tudi predstavitvi uporabljenih algoritmov na praktičnih primerih procesiranja naravnega jezika. **Marjorie McShane** (Univerza v Marylandu, Baltimore County) je na predavanjih skušala odgovoriti na vprašanje, katere jezikovne informacije potrebuje računalniški program za pametnega robota. Skladno s spoznanjem, da je takšnemu programu treba eksplicitno dati vse informacije o entitetah, ki se nanašajo na zunanji svet, oz. povezati določeno besedilo z modelom sveta, se je osredotočila na zahtevnejše primere računalniškega razreševanja reference, kot so skladsenjska in semantična elipsa, izrazi, ki se nanašajo na cele besedilne verige, t. i. premostitveni konteksti, fragmenti stavkov, razrešitev entitete v povezavi z že uskladiščenim védenjem, medjezikovna neujemanja pri pripisovanju reference in razrešitev reference na osnovi skriptov, jih opisala v besedilno-kognitivnem kontekstu ter predstavila algoritme, ki bi jih bilo mogoče uporabiti pri tovrstnem zahtevnejšem razreševanju reference.

V okviru poletne šole so bila organizirana tudi štiri večerna predavanja, večinoma s področij povezave semantike z računalniškim jezikoslovjem in logiko, ki so bila namenjena vsem udeležencem. Prvo je izvedel **Alexander Koller** (Univerza v Kolumbiji), in sicer je predstavil stanje računalniške semantike naravnih jezikov v 21. stoletju. Poudaril je, da se pri gradnji sistemov s pomočjo semantičnih algoritmov upo-

rabljajo spoznanja formalne semantike, da obstajajo učinkovite metode računalniškega razdvoumljanja in logični mehanizmi za ustvarjanje inferenc (t. i. *theorem provers*), da se kaže napredek oz. premik k dejanski rabi jezika, vendar še ni računalniških sistemov, ki bi znali izpeljevati splošne semantične reprezentacije na osnovi dejanskih pomenov besede v kontekstu, manjka pa tudi dovolj obsežna in široka baza védenja kot pomoč pri računalniškem ustvarjanju inferenc, kar se skuša premostiti z učenjem oz. pridobivanjem védenja iz korpusov ter vgrajevanjem semantičnih informacij iz dejanskega sveta. V okviru drugega večernega predavanja je **Michael Witbrock** (Cycorp, Inc.) predstavil bazo védenja Cyc, ki temelji na logiki, poleg tega pa vsebuje tudi mehanizem za računalniško ustvarjanje inferenc na osnovi obsežne baze védenja kot približka umetni inteligenci. **Thomas Ede Zimmermann** (Univerza v Frankfurtu) je večerno predavanje namenil angleškim ustrezniciam glagolov, kot so »iskati«, »lastiti si«, »biti podoben«, katerih predmetna referenca naj bi bila nejasna oz. dvoumna (*opaque verbs*), in angleškim ustrezniciam t. i. upodabljalnih glagolov, kot sta »risati«, »slikati« (*verbs od depiction*), pri čemer se je posvetil predvsem semantičnim reprezentacijam tovrstnih glagolov. **Ronan Reilly** (NUI Maynooth) je na predavanju predstavil nove ugotovitve na področju računalniškega modeliranja jezikovne rabe. Zagovarjal je trditev, da naj bi se dalo slovnico naučiti na podlagi njenih statističnih značilnosti z uporabo SRN (*Simple Recurrent Network*), ki naj bi zadoščale

za obvladanje skladiškovskih struktur in jezikovnih kategorij.

19. poletna šola logike, jezika in informacij je, kar se tiče interdisciplinarnega sodelovanja med jezikoslovjem in računalniškimi vedami, idejno in na praktičnih primerih prinesla veliko novih informacij o t. i. leksikalnih bazah (ontologijah) in bazah védenja na osnovi dejanske rabe jezika v kontekstu v povezavi s semantiko, kognitivnimi vedami in logiko ter o možnostih in načinih uporabe tovrstnih informacij ter gradnje računalniških aplikacij, ki naj bi delovale na osnovi tovrstnih informacij. V slovenskem prostoru je to področje korak, ki do neke mere že kroji delo korpusnih jezikoslovcev (npr. prilagoditev WordNeta slovenskemu jeziku, spletni govorec Klepec), vendar še ne na ravni, ki je bila predstavljena na poletni šoli. Tovrstna raven bo najverjetneje predmet nadaljnjega in prihajajočega dela korpusnih jezikoslovcev.

Jubilejna, 20. poletna šola logike, jezika in informacij, na kateri bodo od 4. do 15. avgusta 2008 predstavljena nova spoznanja na področju interdisciplinarnega sodelovanja med omenjenimi tremi disciplinami, bo potekala v Hamburgu. Vodja programskega odbora je **Rineke Verbrugge** (Univerza v Groningenu), organizacijo pa sta prevzela **Walther von Hahn** in **Cristina Vertan** (Univerza v Hamburgu).

Urška Jarnovič
Univerza v Ljubljani, Filozofska fakulteta
urska.jarnovic@ff.uni-lj.si