

PETA POLETNA DELAVNICA ZA LEKSIKOGRAFIJO IN LEKSIKALNO
RAČUNALNIŠTVO, UNIVERZA MASARYK, BRNO, ČEŠKA REPUBLIKA,
10–14 JUNIJA 2005

Že petič zapored je potekala poletna delavnica za leksikografijo in leksikalno računalništvo *Lexicom*, ki so jo vodile vodilne osebnosti sodobne leksikografije: Sue Atkins, Adam Kilgarriff in Michael Rundell – tj. trojica, ki prireja takšne delavnice v okviru svojega podjetja *Lexicography MasterClass Ltd*. Tokrat je bil *Lexicom* organiziran v sodelovanju s Fakulteto za informatiko češke Univerze Masaryk v Brnu, pod vodstvom Karla Pale.

Udeležencev nas je bilo letos okoli 40, iz Slovenije poleg avtorice poročila še Nataša Logar iz Fakultete za družbene vede Univerze v Ljubljani. Sicer pa je bila skorajda polovica udeležencev iz slovansko govorečih držav – Češka, Rusija, Slovenija, Hrvaška, Srbija in Črna gora, Poljska, Slovaška – ostali pa iz Velike Britanije, ZDA, Južne Afrike, Norveške, Holandije, Francije, Hong-Konga, Japonske in Estonije.

Delavnica je bila zasnovana zelo učinkovito, saj smo udeleženci že nekaj tednov vnaprej dobili predtečajne materiale, ki so bili kot uvod v področje, medtem kot je petdnevni tečaj zajemal poučevanje nadaljevalnih vsebin ter vaje, s čimer je pokrili veliki del teoretične in praktične korpusne leksikografije.

Predtečajni material je vseboval sklope: leksikografske osnove, ki pokrivajo proces izdelave slovarja, opis komponent slovarskih gesel, določanje stila slovarja in njegovih oznak; kako zbirati podatke za slovarje vključno s korpusom kot virom; označevanje korpusov po standardu XML; procesiranje korpusov v leksikografske namene, in sicer označevanje pojavnici, povedi, lem in oblikoslovnih oznak ter skladiščno analizo.

Vsebinsko tečaja pa lahko razdelimo v nekoliko tematskih blokov, ki jih bomo predstavili v nadaljevanju. Osnovna vprašanja kot so pomen besed, pisanje slovarskih gesel in definiranje pomenov so bila teoretično in leksikografsko predstavljena. Kilgarriff nam je govoril o problemih dvoumnosti in nejasnosti, o različnih testih, ki se uporabljajo za medsebojno ločevanje pomenov neke besede ali besedne zveze, o pomanjkljivostih teh testov ter o korpusnemu modelu za določanje pomenov, ki je zasnovan na uporabi besednih skupin (cluster analysis). Rundell je podal pregled tradicionalnih in modernih načinov pisanja definicij v slovarjih ter nam predstavil principe dobrih definicij; neformalno sekcijo pa je tudi posvetil primerjavi in kritiki slovarskih gesel ene besede v različnih slovarjih. Atkinsova nam je govorila o izbiri slovarskih gesel s posebnim poudarkom na večbesednih iztočnicah ter o tem, kako leksikograf v korpusu najde zanj relevantne podatke.

Rundell in Kilgarriff sta nam predstavila načrtovanje in gradnjo leksikografskega korpusa oz. korpusa, ki se lahko uporablja v leksikografiji. Tu sta zaobjela več tem – npr. velikost korpusa, tipi in razmerja vsebovanih besedil, problemi njihovega zbiranja in označevanja ter Kilgarriffovo »poslastico« – tj. teoretične in praktične vidike uporabe spleta kot korpusa. Rundell nas je opozoril na ne preveč ustrezen oz. perspektiven termin »reprezentativen korpus«, ki bi ga raje zamenjal s terminom »uravnotežen korpus«.

Kot izredno uporaben vir za izdelavo slovarjev priporočata Atkins in Rundell slovarske baze z velikimi količinami podatkov. Vsebovale naj bi vse leksikografsko relevantne podatke zajete iz korpusa, nakar se lahko takšne baze uporabijo za izdelavo različnih slovarjev, npr. za enojezične in dvojezične, za šolske, žepne, itd. V predavanju o dvojezični leksikografiji, nam je Atkins na primeru pokazala proces uporabe gesla iz slovarske baze za izdelavo dvojezičnega geselskega članka.

O teoriji semantičnih okvirov (teorija se sicer uporablja v projektu FrameNet) in njeni uporabi v leksikografiji nam je govorila Atkinsova. Na primeru glagola *crawl* je predstavila razlike v pomenih in različne semantične okvirje v katere stopa beseda, ter različne vzorce vezave te besede.

Pred opisom svojega orodja *Word Sketch* nam je Kilgarriff podal pregled korpusne leksikografije od predračunalniškega obdobja, ko se je delalo s kataložnimi listki, prek konkordančnikov in nekoliko kasneje statistično zasnovanih kolokacijskih seznamov do t. i. »besednih skic« (*word sketches*), iz korpusa izhajajočih enostranskih povzetkov slovničnega in kolokacijskega opisa besede. V sodelovanju s češko skupino so razvili spletni program *Word Sketch Engine* (<http://sketchengine.co.uk>), ki ima za izhodišče korpus poljubnega jezika z določenim nivojem jezikoslovnih oznak. Program poleg orodja *Word Sketch* ponuja še konkordančnik, avtomatsko izdelan tezaver in razlike med sinonimi (*sketch difference*).

Udeleženci smo imeli vsak dan nekaj ur namenjenih za praktično delo, kjer smo nad korpusom BNC uporabili *Word Sketch Engine*. V skupinah po trije ljudje smo izdelali par gesel za enojezične in dvojezične slovarje, ter jih predstavili in debatirali o problemih in nejasnostih pri njihovi izdelavi.

Trije predavatelji so si vsebino delavnice medsebojno razdelili in so izmenično predavali, pri čemer so delovali kot dobro uigrana skupina, ki je ne samo med predavanji, temveč tudi med premori ustvarjala izredno pozitivno vzdušje. Sue Atkins, ki je sodelovala s Fillmorjem na izdelavi teorije semantičnih okvirjev in ima bogate izkušnje na področju dvojezične leksikografije, je na predavanjih večinoma predstavljala ta vidik izdelave slovarjev. Adam Kilgarriff, tvorec orodja *Word Sketches*, se je osredinil na računalniško plat korpusne leksikografije, na uporabo spleta kot korpusa, ter določene teoretične probleme. Michael Rundell je kot glavni urednik več slovarjev pokrival v glavnem tematiko, vezano na teoretična in praktična vprašanja leksikografije ter rabo korpusov v leksikografiji. Treba je omeniti še eno ime sodobne korpusne leksikografije – to je Patrick Hanks, ki so ga na delavnici velikokrat omenjali, njegove objave pa največkrat navajali tudi v priročniku tečaja.

Na delavnici smo poslušali tudi štiri gostujoče predavateljce. Gilles-Maurice de Schryver in Phillipe Climent sta predstavila vsak svoj program za sestavljanje slovarjev, *TshwaneLex* ter *DPS* podjetja *IDM*. Kevin Mark je govoril o korpusu tujega jezika *TRIO*, ki je narejen za japonske učence angleščine. Karel Pala, tudi organizator delavnice, je predstavil orodje *DEB* (*Dictionary Editor and Browser*), ki ga uporabljajo v sodelovanju z Inštitutom za češki jezik za izdelavo slovarja toponimov in dialektov, načrtujejo pa njegovo uporabo tudi za izdelavo novega češkega slovarja.

Kot darilo smo udeleženci dobili angleški slovar *Macmillan* (glavni urednik M. Rundell), dobitnika več nagrad za inovativnost. To je prvi slovar narejen z uporabo tehnologij predstavljenih v delavnici (orodje *Word Sketch*), med drugimi novostmi, ki jih prinaša, pa je tudi posebna obravnava najbolj pogostih 7.500 besed, ki so skrbno izbrane po merilih pogostnosti in pomembnosti. Natisnjene so v rdeči barvi ter opremljene z dodatnimi informacijami o njihovih uporabah. Slovar ima tudi CD-rom.

Predstavljena delavnica je lahko še ena spodbuda, da bi se tudi v slovenskem prostoru v prihodnje leksikografsko delo lahko še hitreje premikalo v smeri uporabe korpusov in sodobnih računalniških orodij. K temu bi mogoče pripomogli češki kolegi, njihova pripravljenost za sodelovanje in njihove izkušnje na tekočih projektih izdelave češke leksikalne baze v orodju *DEB*, programa *Visual Browser* za grafično predstavitev leksikalnih odnosov v semantičnem slovarju *WordNet*, spletna predstavitev čeških glagolov z njihovimi valenčnimi okviri *VerbaLex*, sodelovanje pri izdelavi programa *Word Sketch Engine*, ter češka verzija orodja *Word Sketch* s prilagojenimi slovničnimi odnosi za češčino.

Naslednji veliki dogodek, ki ga leksikografi ne bi smeli zamuditi, je 12. mednarodni kongres EURALEX 2006 v Torinu od 6. do 9. septembra 2006 (www.euralex.org). Za marljivejše in prizadevnejše naj še omenimo, da je rok za oddajo člankov 30. oktober 2005.

Pa še to: če koga zanima 253 strani predtečajnega in tečajnega materiala ali nekaj lepih fotografij z delavnice *Lexicom*, jih z veseljem posodim.

Irena Srdanović Erjavec
Filozofska fakulteta v Ljubljani