

Tarptautinė mokslinė konferencija *Presenting terminology and knowledge engineering resources online: models and challenges*

JOLANTA GAIVENYTĖ-BUTLER

Lietuvių kalbos institutas

2010 m. rugpjūčio 12–13 dienomis Dublino miesto universitete (Dublin City University) vyko tarptautinė mokslinė konferencija TKE 2010 *Presenting Terminology and Knowledge Engineering Resources Online: Models and Challenges* (*Terminologijos ir žinių inžinerijos išteklių pateikimas internete: modeliai ir iššūkiai*). TKE konferencijas (International Conference of Terminology and Knowledge Engineering – Tarptautinė terminologijos ir žinių inžinerijos konferencija) kartu su kurios nors Europos šalies universitetu rengia GTW (Gesellschaft für Terminologie und Wissenstransfer – Terminologijos ir žinių perdavimo asociacija). Į šią 1986 m. Vokietijoje įkurtą asociaciją susibūrusios įstaigos, kurių veikla susijusi su kompiuterizuotu terminijos ir žinių kūrimu, tobulinimu ir sklaida. Devintąją TKE konferenciją GTW rengė kartu su Dublino miesto universiteto padaliniu Fiontar. Fiontar – DCU mokymo ir mokslinių tyrimų airių (geilų) kalba padalinys, įsteigtas 1993 m. Čia kuriami elektroniniai airių kalbos ištekliai – tekstynai, interneto priemonės, nacionalinė terminologinių duomenų bazė *Focal*, nacionalinė vietovardžių duomenų bazė *Logainm* ir biografijų airių kalba duomenų bazė *Ainm*.

Konferencijoje dalyvavo pranešėjai iš Airijos, Austrijos, Belgijos, Ispanijos, Jungtinės Karalystės, Kanados, Latvijos, Lietuvos, Meksikos, Prancūzijos, Rumunijos, Suomijos, Švedijos, Vokietijos. Konferencijos dalyvius pasveikino DCU prezidentas prof. Brianas MacCraithas, Fiontar direktorė dr. Caoilfhionna Nic Pháidín ir konferencijos rengėjos Fionnuala de Barra Cusak ir Úna Bhreathnach.

Pirmą konferencijos dieną darbas vyko šešiose sekcijose, skirtose žinių gavimui, kodavimui ir leksinei informacijai, apibrėžtims, terminologijos planavimui ir jos ištekliams, švietimo ir teisės terminų kūrimui ir terminijos tvarkybai (terminology management).

Pilar León Araúz ir Arianne Reimerink iš Granados universiteto (Ispanija) pranešime *Knowledge Extraction and Multidimensionality in the Environmental Domain* (*Žinių gavimas ir daugiamatiškumas aplinkos apsaugos srityje*) kalbėta apie tekstyno modelių analizės (Corpus Pattern Analysis – CPA) taikymą nagrinėjant kontekstinį daugiamatiškumą ir tobulinant patikimo žinių modelių (knowledge patterns – KPs) nustatymo kriterijus terminologinių žinių bazėje EcoLexicon.

Marie-Claude L’Homme, Patricko Leroyer’o ir Benoît Robichaud’o, atstovavusių Monrealio universitetui (Kanada) ir Orhuso verslo mokyklai (Danija), pranešime *Advanced Encoding for Multilingual Access in a Terminological Data Base: A Matter of Balance* (*Progresyvesnis terminologinės duomenų bazės daugiakalbės prieigos kodavimas: pusiausvyros klausimas*) pristatytos naujos, terminologinių duomenų bazėje įdiegtos funkcinės galimybės, palengvinančios prieigą prie daugiakalbių duomenų ir leidžiančios efektyviau juos redaguoti. Vartotojas gali rasti ne tik pagrindinių žodžių, bet ir su jais semantiškai susijusių terminų atitikmenis. Pranešėjų pasiūlyti metodai remiasi formaliu leksinių funkcijų kodavimu.

Bendrame Hammado Afzalo (Nacionalinis gamtos mokslų ir technologijos universitetas, Pakistanas), Paulo Buitelaaro (Nacionalinis Airijos universitetas) ir Philippo Cimiano (Bylefeldo universitetas, Vokietija) pranešime *Generating Lexical Information for Terminology in a Bioinformatics Ontology* (*Leksinė informacija terminologijai bioinformatikos ontologijoje*) pasidalyta pusiau automatinio bioinformatikos ontologijos leksikalizavimo patirtimi taikant LexInfo modelį.

Michalas Boleslavas Mėchura iš Fiontar kalbėjo apie sunkumus nustatant angliškų daugiažodžių terminų reikšmę ir jų atitikmenis kitomis kalbomis. Jo pranešime *When Definitions are not Enough* (*Kai apibrėžčių nepakanka*) išsakyta mintis, kad apibrėžtys, nors ir labai svarbios terminologijoje, nepakankamai detalios atskleidžia vidinę daugiažodžio termino sandarą. Neaiški tokio termino sandara apsunkina kitų kalbų atitikmenų parinktį. Šiai užduočiai palengvinti pranešėjas pasiūlė į sintaksinį medį panašią kompozicinę termino diagramą (compositional term diagram – CTD).

María Pozzi iš El Colegio de Mexico parengė pranešimą *Communicational Model for the Terminological Definition* (*Termino apibrėžties komunikacinis modelis*).

Abolfazlas Zarnikhi (Pompeu Fabra universitetas, Ispanija) pranešime *Towards a Model for Terminology Planning* (*Apie terminologijos planavimo modelį*) kėlė klausimus, kaip ir kodėl terminologijos planavimas skirting-

gose organizacijose remiasi tais pačiais principais, bet nustato savas taisykles, kiek tos taisyklės gali būti bendros, kaip tvirtai terminologijos planavimo veikla remiasi terminologijos teorija, ar galima sudaryti terminologijos planavimo modelį. Pranešėjas lygino Katalonijos terminologijos centro TermCat ir Persų kalbos ir literatūros akademijos terminologijos veiklą ir pasiūlė terminologijos planavimo modelį.

Juditha Kast-Aigner iš Vienos ekonomikos ir verslo universiteto pranešime *Developing and Using Terminology Resources: The Case of EU Terminology (Terminologijos išteklių plėtojimas ir naudojimas: ES terminijos atvejis)* pristatė naudojantis tekstynais atliktą diachroninę Europos Sąjungos dokumentuose vartotos vystomojo bendradarbiavimo (development cooperation) terminijos analizę, atskleidžiančią šios srities sąvokų ir terminų pokyčius.

Albinos Aukšoriūtės ir Jolantos Gaivenytės-Butler (Lietuvių kalbos institutas) pranešime *The Electronic Resources of Lithuanian Terminology (Elektroniniai lietuvių terminijos ištekliai)* buvo pristatyti svarbiausi lietuvių terminijos ištekliai, prieinami internete. Daugiausiai kalbėta apie Lietuvos Respublikos Terminų banką ir Lietuvių kalbos terminyną, jų sandarą, paieškos galimybes, privalumus ir trūkumus.

Fionnualos Cloke ir Gearóido Ó Cleircíno, atstovavusių konferencijos šeimininkams – Fiontar, pranešimas *Irish Language Terms for Legal Translation (Teisės tekstų vertimų airiški terminai)* buvo skirtas teisės aktuose vartojamos airiškos terminijos išteklių kūrimui ir plėtrai apžvelgti. Pristatyto projekto tikslas – teisės aktuose vartojamą airišką terminiją padaryti prieinamą internetu plačiajai visuomenei ir ypač vertėjams. Pranešėjai aptarė keturis jo etapus (lygiagretaus tekstyno kūrimą, terminų išrinkimą, airiškų terminų ir dvikalbio konkordanso dalių atpažinimą, terminų straipsnių redagavimą), pristatė naudotus metodus ir priemones. Parengti terminai perkeliama į Airijos nacionalinę terminologinių duomenų bazę *focal.ie*.

Tegau Andrews iš Bangoro universiteto (Jungtinė Karalystė) supažindino su pastangomis skatinti valų kalbos vartojimą aukštajame moksle ir profesinėje veikloje, kurti ir norminti vališką terminiją. Jos su bendraautorais parengtame pranešime *Developing Welsh Language Terminology in the Higher Education Sector in Wales (Vališkos terminijos kūrimas ir plėtra Velso aukštojo mokslo sektoriuje)* pristatyta taikomoji terminologijos programa *Maes T* ir rengiami internetiniai terminų žodynai.

Estelle Delpech (Lingua et Machina) ir Béatrice Daille (Nanto universitetas, Prancūzija) pranešime *Dealing with Lexicon Acquired from Comparable Corpora: Validation and Exchange (Darbas su žodynu, gautu iš paly-*

ginamųjų tekstynų: patikra ir mainai) kalbėta apie problemas, kylančias dirbant su žodynais, automatiškai gautais iš palyginamųjų tekstynų. Pristatyta specialiai tokiems žodynams sukurta terminų patikros priemonė, duomenų mainams pasiūlytas TBX formatas.

Henrikas Nilssonas iš Terminologicentrum (TNC) (Švedija) pasidalijo Švedijos nacionalinio terminų banko Rikstermbanken kūrimo patirtimi, kalbėjo apie bandymus pritaikyti įvairių tipų žodynus vieningam terminų banko formatui, apie terminų bankų sandaros principus ir galimą jų nesutikimą su vartotojui patogiu terminų pateikimo formatu. Jo pranešime *Layout on Different Terms? (Skirtingas išdėstymas?)* pateikta pavyzdžių iš įvairių terminų bankų ir terminijos tvarkybos sistemų.

Pirmą konferencijos dieną užbaigė kviestinio kalbėtojo Gerhardo Budino iš Vienos universiteto pranešimas *Cognitive Aspects of Designing, Generating, and Using Domain Ontologies (Kognityviniai sritinių ontologijų projektavimo, kūrimo ir naudojimo aspektai)*.

Antra konferencijos diena prasidėjo Fiontar atstovų Michalo Boleslavo Mėchuros ir Briano Ó Raghallaigho pranešimu *How to Build a Termbase for 500,000 Users (and Live to Tell the Story) (Kaip sukurti terminų bazę 500 000 vartotojų (ir išgyventi))*. Pranešėjai supažindino su 2006 m. pradėjusia veikti ir labai didelio populiarumo sulaukusia Airijos nacionaline terminologinių duomenų baze *focal.ie*. Pranešėjų manymu, ši bazė tokia populiari todėl, kad yra patogi vartotojui (lengva paieška, terminų pateikimas labiau primena žodyną negu duomenų bazę).

Vėliau pranešimai buvo skaitomi keturiose sekcijose, skirtose terminų išskyrimui (term extraction), terminologijos mokymui, terminologijos ištekliams ir žinių pateikimui.

Dirko De Hertogo, Kriso Heyleno, Dirko Speelmano (visi iš Leveno universiteto, Belgija) ir Hendriko Kockaerto (Lessius universitetinis katedras, Belgija) pranešimas *A Variational Linguistic Approach to Term Extraction (Variacinės lingvistikos požiūris į terminų išskyrimą)* buvo skirtas automatiškai išrenkant terminus naudojamai pastovaus leksinio žymens analizei (Stable Lexical Marker Analysis).

Patricko Drouino ir Frédérico Dollio (OLST, Monrealio universitetas) pranešime *Corpus Genres and Contrastive Automatic Term Extraction (Tekstyno žanrai ir gretinamasis automatinis terminų išskyrimas)* buvo kalbėta apie tekstyno žanro įtaką automatinio terminų išskyrimo rezultatams. Pranešėjai pristatė žodžių terminiškumui nustatyti taikytą gretinamojo automatinio terminų išskyrimo (Contrastive Automatic Term Extraction -

CATE) metodą, kai lyginamas žodžio dažnis dviejuose skirtingo pobūdžio tekstynuose (bendrajame ir specializuotame).

Georgeta Ciobanu (Timișoaros politechnikos universitetas, Rumunija) skaitė pranešimą *ARGs in Teaching Terminology (ARGs mokant terminologijos)*, kuriame pasidalijo mintimis apie tai, kaip kintamosios realybės žaidimus (ARGs) galima panaudoti moksleivius ir studentus supažindinant su terminologija ir mokant terminijos, ir pristatė Web 2.0 paremtą mokomąjį žaidimą *theSaviours*.

Terminologijos mokymui buvo skirtas dar vienas pranešimas. Anna Aguilar-Amat, Bartolomé Mesa-Lao, Marta Pahisa-Solé ir Pilar Sánchez-Gijón iš Barselonos autonominio universiteto pranešime *BACUS (Base de Coneixement Universitari – University Knowledge Base): 14 Years on Terminology Applied to Translation (BACUS: 14 metų vertimui naudojamos terminologijos)* kalbėjo apie vertimą studijuojančių studentų tyrimų rezultatų pateikimą žinių bazėje BACUS. Ši bazė taip pat gali būti naudojama vertimui, sąvokų, ontologiniams ir tekstynų tyrimams.

Kriso Heyleno (Leveno universitetas, Belgija) ir Hendriko Kockaerto (Lessius universitetinis koledžas, Belgija) pranešimas *The TermWise Knowledge Platform: an Efficient Translation and Terminology Management Suite for Legal Translation in Belgium (TermWise Knowledge Platform: efektyvus Belgijos teisės tekstų vertimo ir terminijos tvarkybos programų kompleksas)* buvo skirtas pristatyti tarpdisciplininiam projektui *TermWise Knowledge Platform*, skirtam teisės tekstų vertimams. Pranešėjai kalbėjo apie dvikalbių terminologinių žinių gavimo iš tekstynų etapus, supažindino su naudotojo sąsaja, jai keliamais reikalavimais ir numatomomis galimybėmis.

Dar vieną terminologijos išteklių – terminologinių žinių bazę *EcoLexicon* pranešime *EcoLexicon: Contextualizing an Environmental Ontology (EcoLexicon: aplinkos apsaugos ontologijos kontekstualizavimas)* pristatė Pilar León Araúz ir Pedro Javier Magaña Redondo iš Granados universiteto.

Andrejaus Vasiljevo (Tilde, Latvijos universitetas), Signės Rirdancės (Latvijos universitetas) ir Tatjanos Gornostay (Tilde) pranešimas *Reaching the User: Targeted Delivery of Federated Content in Multilingual Termbank (Bandyamas pasiekti vartotoją: tikslinis jungtinio turinio pateikimas daugikalbiame terminų banke)* supažindino su internetinės Eurotermbanko vartotojų apklausos rezultatais. Pabrėžta turinio iš daugelio šaltinių konsoliduoto pateikimo svarba.

Anitos Nuopponen (Vasos universitetas, Suomija) pranešime *Transmission of Knowledge about Concept Relations in Terminological Vocabularies*

(Žinių apie sąvokų ryšius perdavimas terminų žodynuose) pateikti internetinės telefonijos terminų žodyno tyrimo rezultatai. Kalbėta apie tai, kaip terminų žodynuose galima perteikti ir panaudoti informaciją apie ryšius tarp sąvokų ir sąvokų sistemų.

Antrą konferencijos dieną taip pat buvo surengta bendra visų konferencijos dalyvių diskusija apie terminologijos ir žinių inžinerijos išteklių pateikimą internete, kurią vedė Henrikas Nilssonas iš TNC. Keletas konferencijos dalyvių pristatė savo įstaigose sukurtus terminologijos ir žinių duomenų bazes, parodė jų galimybes.

Konferencija baigėsi Eugeno Wüsterio prizo teikimu. Šiuo prizu nuo 1997 m. kas treji metai Tarptautinis terminologijos informacijos centras *Infoterm* kartu su Vienos universiteto Eugeno Wüsterio archyvu ir Austrijos nacionaline biblioteka apdovanoja terminologą, kurio darbai turėjo didelės reikšmės terminologijos ar artimų sričių mokslui, mokymui, praktinių metodų tobulinimui ar tarptautinio bendradarbiavimo skatinimui. Eugeno Wüsterio prizą yra gavę profesoriai Heribertas Pichtas, Christeris Laurénas, Juanas C. Sageris, Amelia de Irazazabal Nerpell, Robertas Dubucas, Maria Teresa Cabré, Alanas K. Melby. 2010-ųjų metų Eugeno Wüsterio prizas buvo įteiktas prof. Sue Ellen Wright iš Kento valstijos universiteto Taikomosios lingvistikos instituto (JAV) ir prof. Klausui-Dirkui Schmitzui iš Kelno taikomųjų mokslų universiteto (Vokietija).

Prieš konferenciją ir po jos vyko praktiniai seminarai, kuriuose buvo kalbama apie terminologijos darbo sertifikavimą, pagrindinio terminologijos darbo standarto ISO 704 *Terminology work – principles and methods* (*Terminologijos darbas. Principai ir metodai*) tobulinimą, terminologijos valdymo modelius, jų taikymą ir pokyčius, ISO sąvokų duomenų bazę (ISO Concept DataBase – ISO/CDB) ir TK 37 duomenų kategorijų registrą ISOCat, terminologijos ir ontologijų sąsajas.

Išleistas konferencijos pranešimų rinkinys *Terminology and Knowledge Engineering Conference 2010: Presenting terminology and knowledge engineering resources online: models and challenges* (red. Úna Bhreathnach ir Fionnuala de Barra-Cusack, ISBN 978-0-9566314-0-4).

Gauta 2007-12-11

Jolanta Gaivenytė-Butler
Lietuvių kalbos institutas
P. Vileišio g. 5, LT-10308 Vilnius
E. paštas jolantagb@lki.lt