Daugiakalbių tezaurų rengimas ir adaptavimas: Lietuvos duomenų archyvo LiDA atvejis

Marija Prokopčik
Vilniaus universiteto Komunikacijos fakulteto
Bibliotekininkystės ir informacijos mokslų instituto docentė daktarė,
VUB direktorė informacinei ir kultūros paveldo veiklai
Vilnius University, Faculty of Communication,
Institute of Librarianship and Information, Assoc. Prof., Doctor
Universiteto g. 3, Vilnius
Tel. (8 5) 239 8607
El. paštas: marija.*prokopcik@mb.vu.lt*

Straipsnyje pristatoma daugiakalbių lingvistinių priemonių – tezaurų rengimo ir naudojimo reikšmė elektroninėje erdvėje. Plečiantis kalbų, kuriomis kuriami intelektualūs produktai, ir šių produktų vartotojų bei jų kultūrinės aplinkos įvairovei, daugiakalbiai tezaurai tampa daugiakalbės ir daugiakultūrės informacijos paieškos priemone. Straipsnyje supažindinama su lietuviškų humanitarinių ir socialinių mokslų (toliau HSM) srities daugiakalbių tezaurų rengimo patirtimi. Analizuojamas dviejų kalbų Lietuvos duomenų archyvo LiDA tezauro rengimo procesas. Pagrindžiamas daugiakalbio žodyno rengimo verčiant ir adaptuojant Jungtinės Karalystės duomenų archyvo tezaurą HASSET metodo pasirinkimas, aptariamos parengto tezauro ypatybės, naudojimo ir palaikymo būdai. Lietuviškos daugiakalbio ir daugiakultūrio ELSST tezauro versijos rengimo metodika pristatoma išanalizavus šio tezauro ypatybes ir rengėjų patirtį bei taikytus bendrus principus: kultūrinį neutralumą, visų kalbų versijų vienodą traktavimą, sąvokų suderinimą, terminų adaptavimą kitai kalbai.

Pagrindiniai žodžiai: *informacijos paieška, daugiakalbiai tezaurai, duomenų archyvai.*

Įvadas

Šiuolaikinės informacinės technologijos lengvina visuotinį informacijos ir duomenų perdavimą, tačiau tai nereiškia, kad lengvėja ir globali komunikacija. Jau įprasta, kad elektroninėje erdvėje bendraujama anglų kalba, dėl to kyla tarpkultūrinio komunikavimo problemų visiems tokio proceso dalyviams nepriklausomai nuo to, ar anglų kalba yra jų gimtoji kalba, ar ne. Į tarptautinę areną veržiasi Azijos kalbos, kita vertus, nereikėtų ignoruoti ir mažesnių Europos kalbų, kuriomis sukuriama nemaža dalis pasaulio intelektinio produkto. Daugiakalbiam komunikavimui užtikrinti būtinos ir specialiai parengtos lingvistinės priemonės – rubrikynai, tezaurai, ontologijos. Jos įgyja vis didesnę reikšmę.

Nors informacijos organizavimas ir paieška pastaraisiais metais patiria esminius pokyčius, kai vis labiau populiarėja visatekstė paieška, metaduomenys ir toliau nepraranda savo reikšmės. Jie svarbūs ir elektroninės informacijos autoriams, ir tokios informacijos ieškantiems vartoto-

jams. Profesionaliai parengti ir tinkamai naudojami metaduomenys užtikrina informacijos prieigą nuosekliai „ženklindami“ jos turinį. Metaduomenys rodo kelią link reikalingos informacijos – visos vienoje vietoje.

Tam tikri metaduomenų standartų[1] elementai skirti informacijos turiniui atskleisti ir prieigai prie šio turinio lengvinti. Labiausiai paplitusi turinio ir semantinių informacijos aspektų formalizavimo priemonė yra kontroliuojamas žodynas – tezauras. Pastaruoju metu vis plačiau imama kalbėti apie taksonomijų ir ontologijų kūrimą ir galimybes, tačiau šiame straipsnyje bus nagrinėjami tik tezaurai.

Tezauras – tai svarbi ir naudinga duomenų organizavimo ir paieškos priemonė. Ji padeda vartotojui, indeksuotojui ar paiešką atliekančiam specialistui pasirinkti tinkamiausią terminą arba terminus ir tokius pasirinktus terminus nuosekliai vartoti.

Svarbiausia tai, kad terminų hierarchinių santykių vaizdavimas leidžia vartotojams plėsti paiešką arba prireikus ją siaurinti. Be to, sinonimai leidžia vartotojams pasirinkti skirtingus terminus vienai sąvokai žymėti ir kartu pasiekti tą patį turinį.

Tai, kad daugeliu atvejų tezaurai didina informacijos paieškos efektyvumą, gerėja paieškos rezultatų tikslumas ir (arba) išsamumas, nekelia abejonių (Sihvonen, Vakkari, 2004; Shiri ir kt., 2002). Kai paieška vykdoma be leksikografinės kontrolės, paieškos rezultatui gali turėti įtakos skirtingai formuluojamos užklausos ir skirtingas paieškai vartojamas sąvokas žyminčių terminų suvokimas. Tezaurų paskirtis yra mažinti šią problemą užtikrinant leksikos kontrolę. Tokie kontroliuojami žodynai – ne vien autorizuotų terminų sąrašai. Vartotojams, kurie nepakankamai gerai išmano atitinkamo duomenų masyvo terminiją, tezaurai lengvina prieigą prie indeksuotų informacijos išteklių masyvo. Visos sąvokos yra struktūruojamos ir organizuojamos naudojant griežtai apibrėžtą ryšių sistemą. Tezauruose fiksuojami ryšiai padeda plėsti paiešką, pasirinkti kelis sąvoką žyminčius ekvivalenčius ar semantiškai susijusius terminus (Tudhope ir kt., 2006).

XXI a. informacinių technologijų tolesnė plėtra ir komunikacijos globalizacija formuoja daugiakalbių lingvistinių informacijos organizavimo priemonių poreikį. Viena vertus, didėja vartotojų, jų kultūrinės aplinkos ir kalbų įvairovė, antra vertus, įvairios vartotojų grupės ieško informacijos, kurios šaltinių įvairovė taip pat nuolat auga. Daugiakalbiai kontroliuojami žodynai, visų pirma tezaurai, tampa daugiakalbės informacijos paieškos priemone. Nors daugiakalbių tezaurų skaičius, jų apimamų sferų įvairovė nuolat didėja, nepaisant šiuos darbus reglamentuojančių tarptautinių ir nacionalinių standartų, besiplečiantis daugiakalbės informacijos organizavimo ir paieškos mastas iškelia naujus klausimus ir siūlo naujus jų sprendimo būdus.

Taigi šio straipsnio objektas – daugiakalbiai tezaurai. Straipsnyje keliamas tikslas išanalizuoti bendrųjų metodologinių principų taikymo rengiant daugiakalbius HSM srities tezaurus Lietuvoje galimybes.

[1] Šiame straipsnyje visų pirma kalbama apie Dublino branduolio metaduomenų iniciatyvą (angl. *Dublin Core Metadata Initiative*) – standartą, kurio esmė yra sukurti paprastą metaduomenų schemą, kuri leistų neprofesionalams patiems aprašyti elektroninius informacijos išteklius (Lietuvoje skaitmeninei interneto informacijai aprašyti buvo patvirtintas LST ISO 1586:2007) ir Duomenų aprašymo iniciatyvą (angl. *Data Documentation Initiative)* – standartą socialinių ir humanitarinių mokslų srities mokslinių tyrimų duomenų turiniui ir struktūrai aprašyti, sudarytam aprašui (metaduomenims) atvaizduoti, išsaugoti ir keisti (Data..., 2010).

Teoriniai svarstymai iliustruojami nagrinėjant Lietuvos HSM duomenų archyvui (LiDA) rengiamus tezaurus.

Straipsnyje iškeltam tikslui pasiekti formuluojami tokie uždaviniai:

- išanalizuoti daugiakalbių tezaurų rengimo metodus;
- atskleisti daugiakalbių tezaurų terminijos suderinamumo prielaidas;
- pateikti LiDA lietuvių–anglų kalbų tezauro tobulinimo galimus sprendimus;
- išanalizuoti daugiakalbio ELSST tezauro rengėjų patirtį ir pateikti lietuviškos versijos rengimo gaires.

Rengiant straipsnį buvo naudojami mokslinės literatūros analizės, informacijos grupavimo, sisteminimo, lyginimo ir apibendrinimo metodai.

Tezaurai elektroninėje erdvėje

Kai kurių autorių teigimu, skaitmeninimo epochoje tezaurai išgyvena renesansą, profesionaliai parengti ir palaikomi kontroliuojami žodynai yra esminė kokybiškų informacinių paslaugų teikimo prielaida, o ne vartotojo autonomijos varžymas (The Thesaurus, 2004).

Elektroninių ir tinkle funkcionuojančių tezaurų nuolat daugėja, o jų rengimo, taikymo ir vertinimo metodai kinta ir prisitaiko prie naujų reikalavimų. Greta autonominių ar savarankiškų tezaurų, nesančių kurios nors konkrečios informacijos paieškos sistemos dalimi, vis daugiau tezaurų yra visiškai integruoti į duomenų bazes ar informacijos paieškos sistemas (Shiri, Revie, 2000). Akivaizdu, kad neįmanoma sukurti tezauro, kuris apimtų visus terminus ir tenkintų universalius poreikius, todėl svarbu atrinkti ir organizuoti tam tikrai mokslo (mokslų) ir (arba) praktinei veiklos sričiai būdingus, plačiausiai naudojamus, aiškiai apibrėžtus terminus (McCulloch, 2005), rengti tezaurus atlikus konkrečios srities analizę ir išsiaiškinus individualių vartotojų poreikius (Nielsen, 2001). Antra vertus, nepaisant autorių ir darbų apie kontroliuojamų kalbos žodynų, skirtų informacijos tvarkybai ir ieškai, rengimą didelės įvairovės (Aitchison ir kt., 2000; The Thesaurus, 2004; Broughton, 2005), tarptautinių ir nacionalinių standartų įvairovės (ISO, 1985; ISO, BSI, 1985; BSI, 1987), kiekvieną kartą susidūrus su konkretaus žodyno rengimu, tenka vėl grįžti prie metodologinių ištakų ir svarstyti, kaip sukurti produktą, kuris tenkina tris pagrindines sąlygas: 1) žodynas tiksliai apibūdina duomenų rinkinį, 2) žodyną paprasta valdyti ir atnaujinti, 3) žodyno rengimas ir naudojimas reikalauja minimalių sąnaudų.

Greta tradicinių tezauro rengimo etapų (esamų atitinkamos srities kontroliuojamų žodynų analizė ir naujo tezauro rengimo poreikio įvertinimas, tezauro dydis, terminų pasirinkimas ir atranka, struktūra ir prasminių ryšių nustatymas bei fiksavimas, standartų taikymas, vertinimas ir testavimas, nuolatinis palaikymas ir atnaujinimas) Emma McCulloch pabrėžia tezauro apimamos srities (sričių) specialistų ir tezauro rengėjų konsultavimosi svarbą ir poreikį sukurti specialią tokio konsultavimosi formalizuotą struktūrą ar procedūrą (McCulloch, 2005). Kiti ne mažiau reikšmingi tezauro rengimo uždaviniai yra tinkamos programinės įrangos pasirinkimas ir vartotojo sąsaja. Vartotojo sąsaja formuojama atsižvelgiant į potencialių vartotojų poreikius (naršymas ir (arba) terminų paieška, tezauro hierarchinės struktūros visiškas atskleidimas ir (arba) specifinių sričių išryškinimas, galimybė naudoti visą

tezauro terminų abėcėlinį sąrašą, įskaitant neteiktinus terminus ir pan.

Be to, autorė akcentuoja, kad nors dažniausiai rengėjai kalba apie naują produktą – tezaurą, mažai tikėtina, kad tai bus iš esmės naujas išteklius – greičiausiai didžioji tokio žodyno dalis parengiama sujungus teiktinus ir plačiai taikomus terminus. Esamų atitinkamos tematikos tezaurų, jų adaptavimo ir pritaikymo konkrečioms reikmėms analizė leidžia iš esmės sumažinti tokių žodynų rengimo ir naudojimo sąnaudas.

Tezauro kūrimas yra tęstinis procesas, nes kiekvienoje mokslo srityje ar kryptyje vyksta pokyčiai: atsiranda nauji terminai ar kinta jais reiškiamų sąvokų ribos. Į tezaurą turi būti nuolat įtraukiami nauji terminai (paprastai taikant indukcinį metodą), pašalinami seni, neteiktini terminai. Nuolatinį palaikymą ir atnaujinimą užtikrina aiškiai apibrėžtos tokio proceso taisyklės ir nustatytos atsakomybės ribos.

Akivaizdu, kad daugiakalbės dalykinės žodinės paieškos organizavimą palengvintų sukurta nuo konkrečios kalbos (kalbų) nepriklausoma žinių bazė, kuri taptų visų prieigos kalbų bendra sąvokų struktūra. Tokia žinių bazė palengvintų paiešką ir faktų suvokimą, tačiau, kaip teisingai pažymima (Jorna, Davies, 2001), tokioje žinių bazėje nekreipiama dėmesio į lingvistinius ir kultūrinius prieigos kalbų skirtumus.

Autorės išskiria tris pagrindinius gero šiuolaikinio tezauro bruožus: 1) daugiakalbystė, kuri lengvintų prieigą prie įvairių kalbų informacijos išteklių ir globalios informacinės visuomenės tarpkultūrinę komunikaciją; 2) semantinė struktūra, kuri, priešingai nei abėcėlinė, leistų užtikrinti visų kalbų atstovavimo lygybę (įprastas sprendimas – vartoti vieną kalbą kaip bazinę formavimo (angl. *filing*) kalbą); be to, semantinė struktūra tiksliau ir aiškiau atskleidžia kiekvienos sąvokos prasmę ir kontekstą; 3) vartotojų suvokimo lengvinimas, paieškos išsamumo ir (arba) tikslumo didinimas pateikiant apibrėžimus, vartojimo pastabas, kai kuriais atvejais trumpus enciklopedinio pobūdžio tekstus (Jorna, Davies, 2001).

Daugiakalbis tezauras nereiškia paprasto kelių vienos kalbos tezaurų jungimo. Daugiakalbio tezauro kiekvienos kalbos versija gali būti naudojama savarankiškai ir nepriklausomai nuo kitų, antra vertus, ji yra susijusi su visomis kitomis kalbinėmis versijomis ir atskirai nuo jų negalėtų egzistuoti.

Michelle Hudon atkreipia dėmesį, kad daugiakalbio tezauro rengimas yra daugiau nei paprasta sąvokų ir terminų atitikmenų paieška. Šiam procesui būdingas tam tikras kultūrinis aspektas, todėl ateityje vietoje daugiakalbių tezaurų galbūt vartosime daugiakultūrio tezauro terminą. Be to, išskiriamas ir politinis tokių tezaurų rengimo aspektas, kai vartojamos skirtingą padėtį užimančios kalbos, pavyzdžiui, anglų ir prancūzų kalbos Kanadoje (Hudon, 1997) ar daugiau ir mažiau paplitusios Europos šalių kalbos.

Tikras daugiakalbis tezauras pateikia lygiaverčius sąvokas žyminčių terminų sąrašus kiekviena kalba. Dar svarbiau tai, kad kiekvienu atveju pateikiama iki galo išplėtota tezauro struktūra – visi prasminiai sinonimijos, pavaldumo, asociaciniai, gretimumo ir pan. santykiai ir ryšiai. Taigi, kad ir kurią tezauro kalbos versiją vartotojas pasirinktų, pateikiamos informacijos kiekis ir kokybė yra vienoda.

Daugiakalbiai tezaurai naudojami kaip indeksavimo ir paieškos priemonė

daugiakalbėse informacijos sistemose ar duomenų bazėse, todėl dokumentus galima indeksuoti viena ar keliomis kalbomis (paties dokumento, informacijos tarnybos, duomenų archyvo ir pan.). Paieška gali vykti kita, pavyzdžiui, vartotojo, kalba, o tezauras šiuo atveju atlieka perėjimo kalbos (angl. *switching language*) funkcijas ir lengvina tarpkalbinę komunikaciją.

Daugiakalbis tezauras labai naudingas ir tada, kai vartotojas ketina pateikti užklausą duomenų bazei, kuri „supranta" tik užsienio kalbą. Toks vartotojas daugiakalbiame tezaure galėtų rasti kontroliuojamus terminus paieškos strategijai formuoti. (Hudon, 1997).

Pagrindinės problemos, kurios, M. Hudon nuomone, tradiciškai siejamos su daugiakalbių žodynų rengimu, yra šios:

- kalbos dirbtinis „plėtimas", kad ji atitiktų kitos (užsienio) kalbos sąvokų struktūrą, dėl kurio tokios struktūros tampa sunkiai atpažįstamos ir suprantamos;
- visos sąvokų struktūros perkėlimas iš vienos kultūros į kitą neatsižvelgiant į jos tinkamumą;
- pažodinis terminų vertimas iš šaltinio kalbos, sukuriant beprasmius darinius vertimo kalba.

Daugiakalbių tezaurų rengėjai susiduria su įvairiomis administracinėmis, semantinėmis ir kalbinėmis, technologinėmis kliūtimis. Kuo siauresnė tezauro apimama sritis, tuo specifiškesnius terminus reikia taikyti, tuo sudėtingiau parengti daugiakalbį tezaurą ir užtikrinti jo palaikymą.

Rengiant daugiakalbius tezaurus paprastai vadovaujamasi vienu iš trijų daugiakalbių tezaurų rengimo ISO standarte nurodytų metodų (ISO, 1985):

1. Esamo vienos kalbos tezauro vertimas į vieną ar kelias naujas kalbas. Šis būdas ypač populiarus visų pirma dėl ekonominių priežasčių. Šio metodo didelis pranašumas tas, kad nereikia kurti tezauro struktūros. Akivaizdu, kad toks rengimo būdas neleidžia vienodai traktuoti visų kalbų. Natūralu, kad šaltinio kalba dominuoja ir tuo būdu sukurtas produktas negali tinkamai atspindėti vertimo kalbos ir kultūros. Kadangi vienakalbis žodynas visada yra šališkas tam tikros kultūros atspindys, ir tiesioginis vertimas gali tapti „kultūrinio imperializmo prielaida".
2. Dviem ar keliomis kalbomis jau egzistuojančių vienakalbių tezaurų derinimas ir jungimas. Nors iš pirmo žvilgsnio šis būdas atrodo patrauklus, tačiau jį taikant susiduriama su rimtomis praktinėmis ir intelektinėmis problemomis. Reikia derinti skirtingas hierarchines tezaurų struktūras, atsižvelgti į nevienodą apimtį, terminų platumą, detalumą, kitus esminius skirtumus. Tikėtina, kad didžiausias ir geriausias tezauras užims dominuojančią poziciją, o kitos kalbos turės prisitaikyti, kad atitiktų dominuojančią terminų struktūrą.
3. Naujo daugiakalbio žodyno kūrimas, nesiremiant jau egzistuojančio tezauro terminais, kai vienu metu rengiamos kelios kalbinės versijos. Be abejonės, šis būdas suteikia didesnių garantijų, kad visos kalbos bus traktuojama vienodai. Kiekviena kalba šiuo atveju gali tapti ir šaltinio, ir vertimo kalba, taip išvengiant vienos kalbos ir kultūros dominavimo.

Lietuvos duomenų archyvo anglų–lietuvių kalbų tezauras: HSM srities dviejų kalbų tezauro bandymas

Vykdant Europos socialinio fondo finansuojamą projektą „Empirinių duomenų ir informacijos HSM tyrimams kaupimas ir valdymas: Lietuvos HSM duomenų archyvas (LiDA)“ buvo sukurtas Lietuvos HSM duomenų archyvas (LiDA)[2]. Pagrindinis projekto tikslas buvo sukurti infrastruktūrą, kuri HSM srities mokslininkams ir tyrėjams užtikrintų laisvą ir atvirą prieigą prie empirinių duomenų, skatintų ir lengvintų šios srities tyrimus, naujų ir autentiškų duomenų rinkimą ir naudojimą, mainus su kitomis Europos ir pasaulio archyvinėmis institucijomis. LiDA kaip nacionalinės aprėpties duomenų institucija ne tik kaupia, organizuoja ir teikia mokslininkams ir kitiems tyrėjams prieigą prie struktūruotų Lietuvos duomenų masyvų, ji skatina ir lengvina Lietuvos mokslinių tyrimų integraciją į Europos mokslinių tyrimų erdvę (angl. *European Research Area*). HSM duomenų archyvo atsiradimas lengvina bendradarbiavimą su tarptautiniais socialinių duomenų bazių tinklais ir asociacijomis: Europos Socialinių mokslų archyvų taryba CESSDA (angl. *Council of European Social Science Data Archives*); Rytų Europos duomenų archyvų tinklu EDAN (angl. *East European Data Archive Network*); Tarptautine socialinių mokslų duomenų organizacijų federacija IFDO (angl. *International Federation for Data Organizations of the Social Sciences*); Tarpuniversitetiniu politinių ir socialinių tyrimų konsorciumu ICPSR (angl. *Inter-university Consortium for Political and Social Research*); Tarptautine socialinių mokslų informacinių sistemų ir technologijų asociacija IASSIST (angl. *International Association for Social Science Information Systems and Technology*).

Šiuolaikinės visuomenės raida keičia HSM srities kalbą ir terminus. Dėl šių pokyčių ir visų gyvenimo sričių globalizacijos būtina kurti priemones, kurios leistų lokalizuoti įvairiomis kalbomis kaupiamą informaciją. Kitose valstybėse ir kitomis kalbomis kuriamos ir kaupiamos informacijos ir duomenų poreikis nuolat auga. Tarptautinių informacijos išteklių kūrimo ir naudojimo sėkmė nemažai priklauso nuo vienodo sąvokų supratimo ir įvardijimo. Daugėjant tarptautinių duomenų bazių, didėja ir daugiakalbių lingvistinių priemonių poreikis.

Kuriant LiDA archyvo pildymo duomenų rinkiniais bei prieigos prie šių duomenų sistemą, buvo parengtas LiDA tezauras. Rengiant LiDA tezaurą ir pasirenkant jo sudarymo metodiką vadovautasi nuostatomis, kad LiDA dokumentų tvarkybai ir ieškai reikalingas specialus HSM terminų tezauras lietuvių kalba, kuris taptų LiDA informacinės sistemos lingvistinio aprūpinimo dalimi, užtikrinančia duomenų archyve saugomų dokumentų apdorojimo nuoseklumą ir kokybišką paiešką.

Atlikus kontroliuojamų žodynų rengimo praktikos Lietuvoje analizę[3], buvo

[2] Lietuvos duomenų archyvas LiDA [interaktyvus] [žiūrėta 2011 m. sausio 23 d.]. Prieiga per internetą: <http//www. lidata.eu>.

[3] Lietuvoje iš esmės nėra originalių tezaurų rengimo patirties, dar sudėtingiau kalbėti apie daugiakalbius žodynus. Pirmas ir kol kas vienintelis daugiakalbis tezauras lietuvių kalba – EUROVOC. Lietuvos bibliotekos, rengdamos atitinkamus kontroliuojamus žodynus dokumentų tvarkybai ir ieškai, kaip antai rubrikynus ar hierarchinės struktūros dokumentų klasifikacijas, taip pat naudojasi kitų nacionalinių ar tarptautinių institucijų

nuspręsta, kad LiDA terminų tezaurui sudaryti geriausia pasinaudoti esamais šios srities tezaurais anglų kalbą ir pasirinktą tinkamiausią bei plačiausiai HSM srityje naudojamą išversti į lietuvių kalbą (Varnienė ir kt., 2008).

Atsižvelgus į anksčiau atliktos HSM srities kontroliuojamų žodynų įvairovę, apimtį, paplitimą, atitiktį ISO standartams, buvo padaryta išvada, kad tinkamiausias LiDA tikslams yra HASSET[4] tezauras. Gavus sutikimą jį naudoti nekomerciniais tikslais, HASSET tapo lietuviško HSM srities tezauro pagrindu.

Renkantis LiDA tezauro rengimo modelį buvo nuspręsta, kad atsižvelgiant į Europos ir pasaulio HSM archyvų šiuolaikinę praktiką reikėtų rengti ne vienos kalbos, o dvikalbį anglų–lietuvių kalbų tezaurą. Išanalizavus ISO 5964 standarte pristatomus daugiakalbių tezaurų kūrimo metodų pranašumus ir trūkumus, Lietuvoje gyvuojančias kontroliuojamų žodynų rengimo tradicijas ir praktines rengėjų galimybes, buvo pasirinktas metodas, kai verčiamas kuris nors egzistuojantis vienakalbis tezauras. Svarbiausias pranašumas šiuo atveju buvo tai, kad nereikėjo kurti reikšminių žodžių masyvo, formuoti tezauro struktūros.

Kita vertus, rengėjai suvokė, kad būtina realiai įvertinti galimus tokio darbo sunkumus:

- verčiant nesutampa terminais reiškiamų sąvokų ribos, todėl vieną terminą reikia versti keliais kitos kalbos terminais;
- gali skirtis terminais reiškiamų sąvokų loginiai santykiai, hierarchijos lygis ir pan.;
- visai nėra kitakalbio atitikmens;
- terminų kiekis, jų apimtis neatitinka konkrečių sistemos poreikių.

HASSET tezauro vertimo ir pritaikymo LiDA dokumentų tvarkybai darbas parodė, jog būtina spręsti pirmiau minėtas problemas, ir patvirtino prielaidą, kad atsižvelgiant į konkrečius LiDA poreikius, kaupiamų dokumentų specifiką, akademinės bendruomenės lūkesčius, lietuvių kalbos ypatybes HASSET tezauras turi būti ne tik išverstas, bet ir atitinkamai adaptuotas.

LiDA tezauras[5] – tai dviejų kalbų elektroninis kontroliuojamas žodynas. Žodyne visi terminai yra anglų ir lietuvių kalbomis. Naudotojai vartodami lietuviškus ter-

parengtais žodynais. Taigi kontroliuojamų žodynų rengimo Lietuvoje patirtis rodo, kad dažniausiai pasirenkamas atitinkamų tarptautinių žodynų vertimo į lietuvių kalbą ir adaptavimo būdas.

[4] Jungtinės Karalystės duomenų archyvo (United Kingdom Data Archive, toliau UKDA) *Humanitarinių ir socialinių mokslų elektroninis tezauro* (Humanities and Social Science Electronic Thesaurus, toliau HASSET) pirminis variantas paremtas plačiai žinomu UNESCO tezauru, kurį parengė Jeanas Aitchisonas.

HASSET yra daugiadalykis tezauras, jo turinys ir aprėptis atspindi UKDA poreikius ir išteklius. Plačiausiai ir išsamiausiai pristatytos pagrindinės socialinių mokslų sritys: politika, sociologija, ekonomika, švietimas, teisė, nusikalstamumas, demografija, sveikata, užimtumas, technologijos.

Tezaure mažai tikrinių daiktavardžių. Geografiniai pavadinimai vartojami tiek, kiek jie reikalingi indeksavimui.

HASSET sandara ir struktūra atitinka Didžiosios Britanijos standarto (British Standard 5723:1987) ir ISO 2788-1986 (Establishment and development of monolingual thesauri) reikalavimus.

Tezaure fiksuojami įprastiniai santykiai tarp leksinių vienetų (teiktini ir neteiktini terminai, žymima USE/UF), hierarchiniai santykiai (platesni ir siauresni terminai, žymima BT/NT) ir asociaciniai santykiai (žymima RT). Apibrėžiančių žodžių ir pastabų vartojimas HASSET atitinka bendruosius tezaurams keliamus reikalavimus.

UKDA skatina nekomercinį HASSET taikymą su sąlyga, kad visais tezauro reprodukavimo ar adaptavimo atvejais bus nurodyta UKDA autorystė.

[5] LiDA tezauras [interaktyvus] [žiūrėta 2011 m. sausio 25 d.]. Prieiga per internetą: <http://www.lidata.eu/page.php?page=duomenys_tezauras>.

minus gali atlikti angliškų tekstų paiešką, ir atvirkščiai.

Tezauras apima daugelį HSM ir kitų mokslo bei praktinės veiklos sričių. Iš viso jų yra 36: aplinkos apsauga, ekonomika, darbas ir užimtumas, demografija, komunikacijos, laisvalaikis, sportas ir kultūra, lygios galimybės, socialinė apsauga, sociologija, statistika, sveikatos apsauga, šeima ir šeimos ūkis, švietimas ir ugdymas, teisė, žiniasklaida ir kt.

Kai kurios sritys yra pristatytos detaliau, kitos fragmentiškai. Detalumas daugiausia priklauso nuo atliekamų HSM srities tyrimų tematikos ir atitinkamos terminijos poreikių.

Rengiant LiDA tezaurą, buvo atsisakyta daugelio Jungtinės Karalystė teritorijų, miestų, kitų geografinių pavadinimų, taip pat kai kurių Jungtinei Karalystei būdingų daiktų, reiškinių, procesų pavadinimų. Kai kuriose vietose prie lietuviško termino vartotojas gali rasti pastabą (Jungtinė Karalystė), kuri rodo, kad lietuvių kalba tokio atitikmens nėra, todėl pateikta tik bendroji sąvoka ar paaiškinimas.

Pavyzdžiui, Poor laws – Paramos vargšams įstatymai (JK).

Kita vertus, dėl būtinumo papildyti LiDA tezaurą lietuviškais terminais, pavadinimais ir pan., kurie turi turėti atitikmenį anglų kalba, šie vertiniai gali būti ne visai tikslūs.

Pavyzdžiui, Avarinis gyvenamasis fondas – Emergency dwelling fund (LT).

Visais atvejais LiDA tezaure palikta HASSET tezauro deskriptorių straipsnių struktūra ir užfiksuoti tokie santykiai tarp leksinių vienetų, kurie buvo originale.

LiDA tezaure fiksuojami sinonimijos, gimininiai – rūšiniai ir asociaciniai santykiai, kuriems žymėti vartojami įprasti sutrumpinimai UF, USE, BT, RT, NT (Varnienė ir kt., 2008).

Tezauro kūrimas yra tęstinis procesas, nes kiekviena mokslo sritis ar kryptis kinta: atsiranda naujų terminų ar kinta jais reiškiamų sąvokų ribos. Į tezaurą nuolat turi būti įtraukiami nauji terminai, pašalinami seni, neteiktini terminai. Viena vertus, kaupint ir tvarkant Lietuvos HSM duomenų archyvą atsiranda poreikis įtraukti naujus lietuviškus terminus, kita vertus, keičiasi HASSET terminija.

UKDA skatina nuolatinę terminijos analizę ir jos tobulinimą, todėl HASSET tezauras yra sistemingai papildomas naujais terminais, kai kurių terminų atsisakoma, įtraukiami nauji gimininiai, rūšiniai, susiję terminai, ieškoma aiškesnė terminų reikšmė, siekiant parinkti kuo tikslesnius sąvokas žyminčius terminus. Pradinę HASSET versiją sudarė 4200 teiktinų terminų ir 250 hierarchinių struktūrų. Naujausios 2010 m. HASSET versijos analizė parodė, kad buvo gerokai sumažinta tezauro apimtis: atsisakyta 3200, naujų įvesta apie 600 terminų.

Dauguma terminų, kurių neliko naujausioje HASSET versijoje, susiję su nacionaline ir institucine Jungtinės Karalystės specifika. Pašalinti: Jungtinės Karalystės politinių partijų ir valdžios institucijų pavadinimai (Communist Party of Great Britain, Conservative and Unionist Party); švietimo ir mokymo sistemos specifiniai terminai (Plus Examination, Basic Education, Catering Education); socialinio aprūpinimo sistemos terminai (Carer's Allowance, Death Grant, Industrial Disablement); kitų sričių specifiniai terminai (Communes, Consumer Price Index, Food Stamps); geografiniai pavadinimai, kurių didžiausią dalį sudaro Jungtinės Karalystės vietovių, miestelių pavadinimai.

Įvesti nauji terminai (iš viso apie 600) apima tokias sritis: švietimas ir mokymas; teisė ir nusikalstamumas; socialinis aprūpinimas; sveikatos apsauga ir medicina; demografija, šeimos; darbas ir darbuotojai; ekonomika; aplinkos apsauga; kompiuterija; komunikacija ir kt.

Pasirinktas LiDA tezauro rengimo metodas verčiant ir adaptuojant lietuvių kalbai HASSET tezaurą anglų kalba neišvengiamai reiškia, kad lietuviškas variantas yra antrinis dėl to, kad rengėjams tenka derintis prie Jungtinės Karalystės HSM sampratos ir sistemos bei anglų kalbos struktūros. Tokia situacija lemia tolesnę lietuviškos tezauro dalies tvarkymo ir palaikymo procedūrą. Tobulinant LiDA lietuvių–anglų kalbų tezaurą įgyvendinami tokie sprendimai:

1. HASSET apimamos HSM ir kitų mokslo bei praktinės veiklos sritys be atrankos perkeliamos į lietuvišką tezauro versiją.

2. Adaptuotame tezaure paliekama HASSET tezauro deskriptorių straipsnių struktūra ir fiksuojami tokie santykiai tarp leksinių vienetų, kurie buvo originale. Taigi lietuviškame HSM tezaure fiksuojami sinonimijos, gimininiai – rūšiniai ir asociaciniai santykiai.

Esant reikalui deskriptoriaus reikšmė gali būti siaurinama arba paaiškinama. Tuo tikslu naudojamos vartojimo pastabos (angl. *Scope note*), pavyzdžiui, Money supply – Pinigų pasiūla (SN pinigų, cirkuliuojančių šalies ekonomikoje, kiekis); arba apibrėžiantys žodžiai (angl. *qualifiers*), pavyzdžiui, Incontinence – Nelaikymas (medicina).

3. Atsižvelgiant vietinę specifiką, kultūrines tradicijas, nacionalinius ir institucinius poreikius, yra prasminga įtraukti kai kuriuos specifinius lietuviškus terminus (pavyzdžiui, Seimas, Tautinė mokykla, Moksleivių parlamentas ir pan.), bet pašalinti to paties detalumo / specifiškumo lygio Lietuvos poreikių neatitinkančius kitus specifinius terminus (pavyzdžiui: Church of England, Guerrilla Activities, General Scottish Vocational Qualification ir pan.).

4. Didžioji dalis terminų, kurie atitinka iš anksto apibrėžtas temines ir veiklos sritis, verčiami ir (arba) adaptuojami paliekant hierarchinę deskriptoriaus struktūrą ir siekiant rasti lietuviškus anglų kalbos žodžiais įvardytų sąvokų atitikmenis.

5. Jei nėra tikslaus sąvoką reiškiančio termino, gali būti naudojama pastaba (Jungtinė Karalystė, JK), kuri rodo, kad lietuvių kalboje (a) nėra tokios sąvokos apskritai arba (b) sąvokai reikšti vartojamas ne tikslus terminas arba terminų junginys, o tik sąvokos aiškinimas, kurį galima prilyginti vertimo ir (arba) naudojimo pastabai. Pavyzdžiui:

Advanced level examinations – Aukštesniojo lygio egzaminai (JK)

Upper House – Lordų Rūmai (JK)

6. Adaptuojant lietuvių kalbai naujausią HASSET versiją reikėtų pašalinti kiek galint daugiau terminų, kurie žymi sąvokas, nesusijusias su Lietuvos HSM srities tyrimais. Tokie terminai paprastai yra nesunkiai identifikuojami, nes žymi sąvokas, atspindinčias nacionalines, kultūrines, institucines ir pan. ypatybes. Tai partijų, institucijų, specifinių struktūrų ir pan. pavadinimai.

Nurodyti terminai yra pašalinami. Jų nemato tezauro vartotojai, todėl tokie terminai negali būti vartojami dokumentų rinkinių tvarkybai ir paieškai. Antra vertus, pašalinti terminai išlieka darbinėje tezauro versijoje. Todėl esant reikalui juos galima grąžinti. Be to, pašalintų terminų išlaiky-

mas ir identifikavimas darbinėje tezauro versijoje yra naudingas atliekant tezauro leksikos ir struktūros analizę atnaujintose HASSET versijose.

Reikėtų atsižvelgti į tai, kad kartais yra gana sudėtinga nuspręsti, ar konkretus terminas tikrai turėtų būti pašalintas, pavyzdžiui,

Jaunimo gaujos – Youth gangs

Lūšnynai – Slums

Raitųjų takai – Bridleways

7. Lietuvos HSM tyrėjų poreikius atitinkantys terminai gali būti papildomai įtraukiami į tezaurą atsižvelgiant į konkrečių tyrėjų, institucijų, nacionalinius interesus. Todėl, jei reikia, tezaurą galima papildyti lietuviškais terminais, pavadinimais ir pan. Tokie (neoriginalūs) terminai buvo įtraukti rengiant pirmą lietuvių kalbai adaptuotą HASSET variantą. Papildomai būtų tikslinga įtraukti terminus, žyminčius švietimo ir mokslo, sveikatos apsaugos, teisingumo, socialinės apsaugos sistemų sąvokas, tautų, kalbų (tarmių), teritorinių vienetų ir pan. pavadinimus, sąvokas, siejamas su kultūra ir tradicijomis, ir pan.

Tokie naujai įtraukiami terminai privalo turėti atitikmenį anglų kalba, nors šie vertiniai gali būti ne visai tikslūs.

Atsižvelgiant į jau minėtą tezauro rengimo metodiką, kai paliekama originali tezauro struktūra, siūloma nepildyti originalių deskriptorių straipsnių naujais terminais, išskyrus tuos atvejus, kai to reikia Lietuvos specifikai atspindėti.

Daugiakalbio ELSST tezauro lietuviškos versijos rengimas

2010 m. prasidėjo CESSDA reorganizacija į CESSDA-ERIC, t. y. į Europos socialinių mokslų duomenų archyvų tarybą – Europos tyrimų infrastruktūros konsorciumą. Bendros mokslinių tyrimų duomenų infrastruktūros sukūrimas leis naudotis jungtiniais Europos socialinių duomenų ištekliais taikant naujausias naršymo ir duomenų paieškos technologijas. Savitarpio supratimo memorandumą dėl CESSDA-ERIC įsteigimo greta kitų valstybių pasirašė ir Lietuva.

CESSDA-ERIC uždavinys yra užtikrinti Europos socialinių, humanitarinių ir kitų gimininguų disciplinų tyrėjams įmanomai geriausią ir paprasčiausią prieigą prie aukščiausios kokybės duomenų. Pasirašydama memorandumą, Lietuva įsipareigojo prisidėti prie CESSDA-ERIC prioritetų realizavimo, kaip antai dokumentuojant duomenis taikyti vienodus DDI metaduomenų standarto elementus; užtikrinti, kad vartotojai duomenis atsisiųstų per bendrą prieigą; išlaikyti vietines kalbas daugiakalbiuose tezauruose ir kt. (CESSDA, 2010).

CESSDA portale naudojamas daugiakalbis Europos kalbų socialinių mokslų tezauras *(European Language Social Science Thesaurus,* toliau ELSST). ELSST rengimo tikslas – sukurti CESSDA narių duomenų indeksavimo ir paieškos priemonę, kuri lengvintų prieigą prie Europos archyvuose kaupiamų duomenų, nepriklausomai nuo jų saugojimo vietos, vartojamos kalbos ir žodyno. Lingvistinių priemonių rengimas yra vienas iš svarbių uždavinių, sudarančių sąlygas ateityje lengviau naudotis CESSDA-ERIC teikiamomis galimybėmis. Kiekviena nauja CESSDA-ERIC narė privalo užtikrinti, kad tapusi nare per devynis mėnesius į savo šalies kalbą išvers visus kontroliuojamų žodynų terminus ir reikšminius žodžius.

Pasirašydama memorandumą, Lietuva įsipareigojo parengti lietuvišką ELSST versiją. Tai leistų Lietuvos tyrėjams prieiti prie įvairių kalbų tyrimų medžiagos ir duomenų, be to, lengviau integruoti lietuviškus

išteklius į bendrą europinę infrastruktūrą. Daugiakalbio ir *daugiakultūrinio* HSM srities tezauro rengimas – svarbi mažesnių ir vietinių kalbų naudojimo ir išlaikymo daugiakalbiuose tezauruose priemonė. Pastarasis aspektas ypač reikšmingas dėl to, kad HSM ir gretutinių sričių daugiakalbių terminologinių priemonių analizė parodė, kad vos keli žodynai turi lietuviškas versijas (žr. lentelę).

Pavadinimas	**Pavadinimas anglų kalba /adresas**	**Kalbos**
UNESCO tezauras	UNESCO thesaurus http://www2.ulcc.ac.uk/unesco/	Anglų, prancūzų, ispanų
EBPO makrotezauras	OECD macrothesaurus http://168.96.200.17/ar/oecd-macroth/	Anglų, prancūzų, ispanų
JT informacinės sistemos tezauras	UNBIS http://lib-thesaurus.un.org/LIB/DHLUNBISThesaurus.nsf	Anglų, prancūzų, ispanų, rusų, kinų, arabų
Bendrasis aplinkos tezauras	GEMET (General Multilingual Environmental Thesaurus) http://www.eionet.europa.eu/gemet/about	Iš viso 28 kalbos (lietuvių)
Sveikatos apsaugos tezauras	European Multilingual Thesaurus on Health Promotion in Twelve Languages http://www.taxonomywarehouse.com/vocabdetails_include.asp?vVocID=%2072	Anglų, prancūzų, ispanų, vokiečių, italų ir kt.
Mokymosi išteklių tezauras	LRE http://www.eun.org/eun.org2/eun/en/etb/content.cfm?lang=en&ov=7208.	Anglų, prancūzų, ispanų, vokiečių, italų, danų, graikų, švedų, albanų, hebrajų
Europos švietimo sistemų tezauras	TESE – Multilingual Thesaurus on education systems in Europe: http://eacea.ec.europa.eu/eurydice/portal/page/portal/Eurydice/TESEHome	Anglų, olandų, čekų, estų, suomių, prancūzų, ispanų, vokiečių, italų, lenkų, portugalų
Europos švietimo tezauras	European Education Thesaurus REDINED http://redined.ar020.com.ar/en/index.php	Anglų, prancūzų, ispanų, italų ir kt. (iš viso 13)
Gyventojų daugiakalbis tezauras	Population Multilingual Thesaurus POPIN http://www.cicred.org/OLD2004/thesaurus/INTEGRAL/index.html	Anglų, prancūzų, ispanų
TDO tezauras	ILO Thesaurus http://www.ilo.org/public/english/support/lib/tools/aboutthes.htm	Anglų, prancūzų, ispanų
Europos profesinio rengimo plėtros centro tezauras	European training Thesaurus http://libserver.cedefop.europa.eau/ett/en/	Anglų, prancūzų, ispanų, vokiečių portugalų, graikų
JT žemės ūkio terminų tezauras	AGROVOC http://aims.fao.org/website/AGROVOC-Thesaurus/sub	Anglų, prancūzų, ispanų, čekų, vengrų, italų, rusų, lenkų
ES tezauras EuroVoc	EUROVOC. The EU's multilingual thesaurus http://europa.eu/eurovoc/	ES kalbos (lietuvių)
Europos terminų bankas	IATE interactive terminology in Europehttp://iate.europa.eu/iatediff/SearchByQueryLoad.do?method=load	ES kalbos (lietuvių)
Europos švietimo terminų žodynas	European Glossary on Education http://eacea.ec.europa.eu/education/eurydice/tools_en.php	Anglų, prancūzų, ispanų, vokiečių, portugalų, italų, rumunų

2010 m. ELSST buvo naudojamos devynių kalbų žodyno versijos: anglų (pagrindas ir šaltinis), danų, graikų, ispanų, norvegų, prancūzų, suomių, švedų ir vokiečių.

ELSST sudaro aukštesnio lygio (platesni) terminai, perimti iš HASSET hierarchinės struktūros. Perėmimo, modifikavimo ir adaptavimo tikslas buvo eliminuoti ir (arba) sumažinti institucinę ir kultūrinę terminijos specifiką.

ELSST yra daugiadalykis tezauras, kurį sudaro šios teminės sritys: ekonomika, darbas ir įsidarbinimas, politika, politinės sistemos, politinės institucijos, sociologija, socialinė struktūra, socialinės problemos, socialinė gerovė, diskriminacija, probleminės grupės, gyvenimo sąlygos, amžiaus grupės, demografija, duomenys, edukologija, tapatybė, tautiškumas, etninės grupės, šeima, šeimos aplinka, religija, požiūriai, analizė, metodologija, aplinkosaugos mokslai.

Šiuo metu ELSST sudaro 3329 deskriptoriai. Iš viso tezaure yra 208 aukščiausi, arba hierarchiniai, terminai. Tezauro turinys yra nuolat peržiūrimas, kai kurių terminų atsisakoma, keičiama teminių hierarchijų struktūra, įtraukiami nauji teiktini terminai, jų sinonimai, vartojimo pastabos ir terminų turinį tikslinantys apibrėžimai. Kartą per metus ELSST tezauras yra atnaujinamas. ELSST valdymo grupės sprendimu vienu metu funkcionuoja trys tezauro versijos[6].

Tezaure išskiriami hierarchiniai (angl. *hierarchical*) ir nehierarchiniai (angl. *non-hierarchical*) santykiai.

Hierarchiniai santykiai ir jų žymėjimas:

- aukščiausias terminas – TT (angl. *top term*),
- platesnis terminas – BT (angl. *broader term*),
- siauresnis terminas – NT (angl. *narrower term*).

Nehierarchiniai santykiai ir jų žymėjimas:

- nevartotini (netinkami) terminai (sinonimai, antonimai, homonimai, skirtingai rašomi žodžiai) – UF (angl. *non-preferred, used for*),
- susiję terminai (asociacijų pagrindu) – RT (angl. *related term*),
- atitikmenys kitomis kalbomis (angl. *multilingual equivalence*).

Tezauro terminų tikslinimas – tai būdas spręsti sąvokas žyminčių terminų nevienareikšmiškumo problemas. Terminams tikslinti taikomos vartojimo pastabos ir apibrėžiantys žodžiai, kurie apibūdina termino prasmę ir vartojimą konkrečioje srityje, gali apriboti termino vartojimo sritį arba įspėti, kad reikėtų vartoti kitą terminą.

---

[6] *Dabartinė versija* (angl. *Present Version*). Tai pastovi šiuo metu naudojama ELSST versija, kuri galioja vienerius metus. Šioje versijoje negalima daryti jokių pakeitimų. Vartotojai gali teikti siūlymus, kurie galėtų būti įtraukti į Būsimą tezauro versiją.

Būsima versija (angl. *Next Version*). Ši tezauro versija atspindi judėjimą link Būsimos tezauro versijos, kuri išleidžiama einamųjų metų pabaigoje. Keitimus šioje versijoje gali daryti tik ELSST administratoriai ir iš dalies vertėjai (jie gali išversti terminus ir pridėti / redaguoti / pašalinti vertimo pastabas (SN) arba sinonimus (UF) savo kalba). Pakeitimai galimi tik tada, kai terminus, kurie buvo įvesti į komentarų duomenų bazę, aprobavo ELSST tezauro valdymo grupė.

Vietinė išplėstinė versija (angl. *Local Extension*). Tai ELSST versija, kuria vietos valdytojai gali naudotis visiškai laisvai: įtraukti naujus terminus ir terminų prasminius ryšius tais atvejais, kai vienas jų yra vietinis (t. y. ne ELSST terminas). Jeigu manoma, kad Vietinės išplėstinės versijos terminas galėtų būti tinkamas tapti ELSST leksiniu vienetu, jis įtraukiamas į Komentarų DB, iš kurios prireikus gali būti perkeltas į Būsimą ELSST versiją. Jeigu nusprendžiama, kad toks terminas nėra tinkamas kaip ELSST leksinis vienetas, jis netraukiamas į Komentarų DB, o lieka Vietinės išplėstinės versijos leksiniu vienetu.

Daugiakalbis CESSDA tezauras ELSST buvo rengiamas etapais, darbai paprastai atliekami vykdant kurį nors europinį projektą (LIMBER, MADIERA). Daugiakalbio tezauro rengėjų vertimo į kitas kalbas patirtis apibendrintai pateikiama keliose publikacijose (Kluck, Huckstorf, 2008; Multilingual, 2006). Ypač nuosekliai šio tezauro vertimo ir adaptavimo patirtį aprašo Suomijos specialistai (Jaskelainen, Forsman, 2003; Keranen, 2002; Nykyry, 2010).

Visų pirma reikėtų pristatyti bendruosius ELSST rengimo principus kaip pagrindą naudojant UKDA tezaurą HASSET, svarbiausius šio darbo etapus, principus ir metodus (European, 2010) . Nors HASSET apibūdinamas kaip socialinių ir humanitarinių mokslų tezauras, atsižvelgiant į socialinės aplinkos įvairovę ir tyrėjų, politikų, žurnalistų poreikį gauti platų visuomenės gyvenimo Europoje vaizdą, socialinių ir humanitarinių mokslų blokas buvo papildytas aplinkos apsaugos, sveikatos priežiūros, geografijos terminais. Pirmame daugiakalbio žodyno rengimo etape (dalyvaujant LIMBER projekte) buvo parengtas keturių kalbų – anglų, ispanų, prancūzų ir vokiečių kalbų tezauras.

Tezauras buvo rengiamas dviem etapais:

1. *Vienos kalbos tezauro apimties mažinimas*

Pasirinkus kaip pagrindą HASSET, kuris buvo laikomas geriausiu socialinių mokslų archyvų organizavimo tezauru (4,2 tūkst. teiktinų terminų ir 250 hierarchijų), buvo suformuota tokia šio bazinio tezauro struktūros keitimo ir pritaikymo bendroms Europos reikmėms strategija:

- 26 pagrindinių hierarchijų (teminių klasių) stambinimas einant nuo plačiausių link specifiškų terminų, kad būtų eliminuota kultūrinė ir institucinė specifika. Kita vertus, toks mažinimas ir vietinės specifikos eliminavimas vyko suvokiant, kad ELSST turėtų tapti bendra ontologija, kurią būtų galima plėtoti atsižvelgiant į konkrečių archyvų institucinius ir kultūrinius poreikius, o esant reikalui net susieti su kitais tam tikrų sričių tezaurais;
- analizė ir hierarchijų stambinimas leido sumažinti teiktinų terminų skaičių iki 2,5 tūkst.;
- iš pradžių ELSST rengėjai atrinko tokias hierarchijas: ekonomika, darbas ir įsidarbinimas, politika, politinės sistemos, socialinės problemos, diskriminacija, požiūriai, probleminės grupės, politinės institucijos, etninės grupės, gyvenimo sąlygos, socialinė struktūra, duomenys, amžiaus grupės, demografija, sociologija, socialinė apsauga, aplinkosaugos mokslai, edukologija, tapatumas, nacionalinis savitumas, šeima, religija, analizė, metodologija, šeimos aplinka;
- nacionalinės ir institucinės specifikos eliminavimas leido sumažinti teiktinų pirmosios versijos ELSST terminų skaičių iki 1 tūkst., todėl vėliau terminų bazė buvo papildyta dar 23 susijusiomis hierarchijomis ir pirmoje sumažinto vienakalbio tezauro versijoje (2001-05-30) buvo 1380 teiktinų terminų (deskriptorių) ir 49 hierarchijos.

2. *ELSST vertimas*

Rinkdamiesi šio daugiakalbio tezauro rengimo metodą, autoriai išnagrinėjo ISO 5964-1985 gairėse apibūdintų daugiakal-

bių tezaurų rengimo metodų pranašumus ir trūkumus.

ELSST tezaurui rengti buvo pasirinktas pirmasis daugiakalbių tezaurų rengimo metodas – esamo tezauro vertimas. Suvokdami šio metodo keliamus pavojus, rengėjai nurodė, kad sąvoka „vertimas“ gali būti interpretuojama skirtingai ir priklauso nuo tokio vertimo tikslo ir to, kaip jis bus naudojamas. Vertimo tikslas gali būti siekis padėti išversto teksto vartotojui geriau suprasti šaltinio kalbos terminus. Šiuo atveju vertimo žodžiai gali netikti vartoti kaip indeksavimo terminai. Kita vertus, jeigu išverstus žodžius ketinama vartoti kaip indeksavimo terminus, jie gali neatitikti (iš dalies atitikti) šaltinio terminų reikšmės. ELSST buvo rengiamas taikant antrąjį metodą, todėl vertimas įgavo tam tikrų daugiakalbių tezaurų rengimo metodo bruožų, kai vienu metu rengiamos kelios kalbinės versijos.

ISO 5964-1985 gairėse nurodoma, kad nepaisant pasirinkto rengimo metodo naujojo tezauro terminai turi turėti tokį patį statusą kaip ir bazinio (verčiamojo, šaltinio kalbos) tezauro. Tai yra ELSST rengėjų tikslas, nes šis tezauras yra ne tik daugiakalbis, bet ir daugiakultūris, todėl teiktini terminai turėtų atspindėti Europos, o ne nacionalinį kontekstą. Kita vertus, kalbos, regiono, institucijos, kultūros specifiką perteikiantys terminai gali būti vartojami kaip atitiktiniai (angl. *non-preferred*).

Atitikties sąvoka ELSST yra taikoma tik terminams, ne hierarchinėms struktūroms, todėl skirtingų kalbų tezaurai gali turėti skirtingas hierarchines struktūras. Kita vertus, terminų atitikties lygis svarstomas tik ryšium su teiktinais terminais, atitiktinių terminų tikslus vertimas ne visada įmanomas ir (arba) reikalingas. Sprendžiant terminų neatitikties problemą siūloma atsižvelgti į vertimo naudojimo pastabas.

Skaitmeninėje aplinkoje skirtingų terminų taikymas vienodoms sąvokoms įvardyti yra viena pagrindinių rezultatyvios paieškos kliūčių. Dėl šios terminologinės problemos vartotojams ne visada pavyksta prieiti prie tarpdalykinių ir daugiakalbių išteklių (McCulloch ir kt., 2005). Taina Jaaskelainen (2006) nurodo, kad svarbiausias daugiakalbių tezaurų rengimo ir terminų vertimo uždavinys yra sąvokų suderinimas (angl. *mapping*).

Martinas Doerris suderinimą apibūdina kaip „beveik ekvivalentiškų terminų, sąvokų ir hierarchinių ryšių atpažinimo procesą“ (Doerr, 2001). Tezaurai naudojami indeksavimui, t. y. informacijos vienetų turinio apibūdinimui ir informacijos paieškai pagal dalyką, todėl jie atlieka savo funkciją tik tada, kai įvairių kalbų atitinkami terminai žymi tą pačią sąvoką. Kitas svarbus reikalavimas – atrinkti terminai yra plačiai ir bendrai vartojami vertimo kalboje arba juos įprasta vartoti profesinėje bendruomenėje. Sąvokų suderinimas užtikrina galimybę rasti duomenų rinkinius apie tą patį dalyką ir (arba) temą įvairiuose skirtingų kalbų archyvuose ir sąvokų skirtingomis kalbomis suderinamumą.

Terminų (žodžių ar žodžių junginių) pasirinkimas sąvokai išreikšti nėra labai svarbus, o pažodinis vertimas kartais gali iškreipti sąvokos reikšmę. Taigi pirmiausia vertėjas turi išsiaiškinti, kokią sąvoką žymi šaltinio terminas, kokia tos sąvokos apimtis ir vartojimas. Tai ne visada paprasta, nes tezaure terminai pateikiami be konteksto. Paprastai įmanoma pasiekti visišką arba dalinę sąvokas žyminčių terminų atitiktį, bet pasitaiko išimčių. Daugiausia problemų kyla dėl kalbinių ir kultūrinių skirtumų,

kai tam tikros sąvokos, kurios egzistuoja bazinėje kalboje, vertimo kalboje apskritai neegzistuoja. Pateikiamas pavyzdys – anglų kalbos sąvoka „blood sports“ (sportas, kai žudomi arba žalojami gyvūnai). Vertėjas turi tris alternatyvas: 1) tiesioginis vertimas, kurio rezultatas dirbtinis terminas, be to, šiuo atveju reikėtų nurodyti visas atitiktines sąvokas (šunų / gaidžių ir pan. kautynės); 2) atitikmuo – artimiausia siauresnė arba platesnė sąvoka; šiuo atveju būtina vartojimo pastaba, apibūdinanti bazinio (angliško) ir vertimo kalbos sąvokų skirtumus; 3) kelių terminų derinys, bandant paaiškinti sąvokos reikšmę. Vertėjai siūlo rinktis antrą variantą ir vartoti terminą „medžioklė“ bei žiūrėti vertimo naudojimo pastabas. Šis terminas yra siauresnis, tačiau bendrai vartojamas.

Didelis iššūkis – švietimo, socialinės apsaugos, sveikatos priežiūros, teisės sistemų skirtumai ir šių skirtumų nulemtos skirtingos sąvokos ir jas žymintys terminai. Autorės nuomone, šiuo atveju geriau vartoti bendrus ar standartinius terminus, paprastai platesnes sąvokas žyminčius terminus. Atitikmenis daug lengviau rasti tada, kai tezauro terminai bus kiek galint mažiau susiję su kultūros specifika ar tradicija.

Apibendrindami ELSST kaip daugiakalbio devynių kalbų tezauro kūrimo patirtį konferencijoje „Categorising Human Knowledge: Classification in Languages and Knowledge Systems“ (COST A31 konferencija, 2010 m. gegužės 14–16 d., Paryžius), žodyno vertėjai ir sudarytojai nurodė, kad daugiakalbio žodyno ypatingas skiriamasis bruožas yra terminų įvairiomis kalbomis pasirinkimas ir jų atitikties laipsnis (European, 2010). Turint omeny, kad kalbų struktūra, terminų tradicinės reikšmės, dalykinių sričių terminijos apimtis ir vartojimas gali skirtis, taip pat atsižvelgiant į kultūros kontekstą, išskiriami keli terminais išreikštų sąvokų atitikties lygiai (ISO, 1985):

- Visiška atitiktis (viena sąvoka visomis kalbomis);
- Nevisiška atitiktis (angl. *inexact/ near equivalence*) – gali įvairuoti priereikšmis;
- Dalinė atitiktis (sąvoka viena kalba platesnė arba siauresnė nei kita);
- Atitiktis „vienas terminas – keli terminai“ (verčiant reikia kelių terminų pirminei sąvokai išreikšti);
- Neatitiktis.

Be to, keliomis kalbomis išreikštų sąvokų atitikties lygiai turi būti suvokiami svarbiausių tezauro vykdomų funkcijų kontekste.

Svarbi HSM terminijos ypatybė yra ta, kad šių mokslų sąvokos ir terminai priklauso „minkštųjų“ (angl. *Soft-core*) kategorijai. Priešingai nei „kietosios“ (angl. *Hard-core*) sąvokos, kurios yra aiškiai apibrėžtos ir kurių apibrėžimą taiko visa suinteresuota bendruomenė, „minkštosios“ yra paprastai abstrakčios, neapibrėžtos, daugiareikšmės. HSM srities „minkštųjų“ sąvokų reikšmei parodyti ir (arba) patikslinti yra svarbus socialinis kontekstas. Pavyzdžiui, sąvokos *senatvė* arba *vaikystė* įvairiose šalyse yra suvokiamos skirtingai, atsižvelgiant į vidutinę gyvenimo trukmę.

Be to, sąvokos reikšmė ir ją nusakantis terminas ilgainiui gali keistis. Pavyzdžiui, terminas *skurdas* yra įtrauktas į tris ELSST hierarchijas: *ekonomika, socialiniai rodikliai ir socialinė atskirtis*.

ELSST rengėjų patirties ir ELSST leksikos bei struktūros analizė leido suformuluoti pagrindinius lietuviško ELSST varianto rengimo principus:

1. Rengiant daugiakalbį tezaurą svarbiausia tinkamai perduoti sąvokos reikšmę, todėl toks darbas tik sąlygiškai gali būti laikomas vertimu, tai greičiau originalo anglų kalba adaptavimas kuriai nors kalbai. Konkrečių terminų (žodžių ar žodžių junginių) pasirinkimas sąvokai išreikšti nėra labai svarbus, o pažodinis vertimas kartais gali iškreipti sąvokos reikšmę. Pavyzdžiui,

Teisė naudotis svetima žeme einančiais takais ar keliais – Rights of way; Rajonai, kuriuose didelis nedarbas – Depressed areas.

2. Dauguma sąvokų, ypač žyminčių platesnes hierarchines klases, ir joms reikšti vartojami terminai visiškai atitinka. Tokių sąvokų yra maždaug 70 proc. Tai svarbus leksinių vienetų masyvas, kuris užtikrina galimybę rasti duomenų rinkinius apie tą patį dalyką ir (arba) temą įvairiuose skirtingas kalbas naudojančiuose archyvuose ir sąvokų skirtingomis kalbomis suderinamumą. Šis leksinių vienetų masyvas svarbus ir tais atvejais, kai specifiškesnės šių hierarchinių klasių sąvokos neturi tikslių atitikmenų vertimo kalba.

3. Nevisiška ir (arba) dalinė sąvokos ir ją žyminčio termino atitiktis paprastai susijusi su leksiniais ir kultūrų skirtumais. Pavyzdžiui, sąvokas Housewives ir House husbands tik iš dalies atitinka žodžių junginiai Namų šeimininkės ir Namų ūkį tvarkantys vyrai, skiriasi jų priereikšmiai. Jei reikia, tokias dalinai ir (arba) ne visai atitinkančias sąvokas galima paaiškinti pasitelkiant vartojimo pastabas arba vertimo vartojimo pastabas. Pavyzdžiui, Nederamas elgesys – Abuse (SN – plati sąvoka, į kurią įeina užgauliojimas, įžeidimas, keiksnojimas ir kt.).

4. Ieškant lietuviškų terminų sąvokoms žymėti, dažnai tenka sąvokas, kurias anglų kalba žymi vienas žodis arba stabilus žodžių junginys, versti keliais žodžiais, fraze. Toks vertimas įgyja ir vartojimo pastabos statusą. Pavyzdžiui, Darbo sutarties normų laikymasis (streiko forma) – Work-to-rule; Protestai, kai užsirakinama pastato viduje – Lock-ins; Kombinuoti mokymai (paskaitų ir gamybinės praktikos derinys) – Sandwich courses.

5. Nepaisant ELSST rengėjų pastangų eliminuoti šalių kultūroms ir (arba) institucijoms būdingus terminus, tokių sąvokų vis dar pasitaiko. Blogiausias sprendimas šioje situacijoje – pažodinis vertimas. Lietuviško varianto rengėjams reikėtų išsiaiškinti tokių sąvokų tikslų turinį ir vartoti platesnį ar siauresnį terminą (pavyzdžiui, *new age travellers – klajokliai, public gardens – parkai*) ir esant reikalui teikti vertimo vartojimo pastabą. Kultūrines ypatybes perteikiančius specifinius terminus reikėtų vartoti kaip atitiktinius, pavyzdžiui: Coroner – Koroneris (SN specialus teismo tardytojas, tiriantis staiga mirusių asmenų mirties priežastis, kai yra pagrindo įtarti, kad mirtis įvyko dėl smurto); Use – Teismo tarnautojai.

6. Visais atvejais, kai sąvokai žymėti vartojami terminai yra nevienareikšmiai, patariama vartoti apibrėžiančius žodžius. Pavyzdžiui, Vaisingumas (Demografija) – Fertility; Vaisingumas (Medicina) – Fecundity.

7. Visais atvejais, kai tenka rinktis ar pažodinį vertimą, ar netikslų, bet paplitusį ir bendrai vartojamą terminą (terminus), pirmenybė turėtų būti teikiama pastariesiems. Pavyzdžiui, Mokslo baigimo pažymėjimai – Educational certificates; Ko-

lektyviniai sodai – Private gardens; Vidaus kiemai – Garden patios.

8. Jeigu manoma, kad ELSST būtina papildyti teiktinu lietuvių kalbos terminu (deskriptoriumi), jis įtraukiamas į specialią komentarų DB (angl. *comments database*). Siūlymų analizė ir aprobavimas vyksta pagal nustatytą procedūrą. Siūlomas terminas turi atitikti dvi sąlygas – sąvokos neapima joks esamas terminas ir sąvoka neturi ryškios kultūrinės ir institucinės priklausomybės. Naujas sąvokas gali siūlyti atskiros institucijos. Siūlomas terminas turi turėti atitikmenį anglų kalba, vartojimo pastabą (nebent būtų labai aiškiai suprantamas kaip toks), nurodoma jo vieta hierarchinėje struktūroje (platesnis arba aukščiausias terminas). Galima siūlyti ir tokio termino sinonimus.

Siūlymų dėl terminų redagavimo, šalinimo, struktūros keitimo gali teikti atskiros institucijos. Tokie siūlymai anglų kalba įtraukiami į specialią komentarų DB. Siūlymus analizuoja ELSST valdyba. Nebevartojami terminai ir toliau lieka tezaure, tačiau jie netaikojami indeksavimui.

Parengtas lietuviškas ELSST variantas turėtų būti testuojamas ir vertinamas naudojant jį duomenų paieškai CESSDA kataloge. Testavimo rezultatai galėtų būti pagrindas tobulinti Būsimą ir plėtoti Vietinę tezauro versijas. Norint užtikrinti lietuviško ELSST varianto kokybę, būtina numatyti tezauro palaikymo procedūrą, kuri leistų kiekvienais metais atnaujinti tezaurą atsižvelgiant į ELSST valdybos priimtus sprendimus.

Išvados

Šiuolaikinės visuomenės raida keičia HSM srities kalbą ir terminus. Dėl šių pokyčių ir visų gyvenimo sričių globalizacijos būtina kurti priemones, kurios leistų lokalizuoti įvairiomis kalbomis kaupiamą informaciją. Kitose valstybėse ir kitomis kalbomis kuriamos ir kaupiamos informacijos ir duomenų poreikis nuolat didėja. Tarptautinių informacijos išteklių kūrimo ir naudojimo sėkmė nemažai priklauso nuo vartojamų sąvokų vienodo supratimo ir įvardijimo. Daugėjant tarptautinių duomenų bazių, auga ir daugiakalbių lingvistinių priemonių poreikis.

Daugiakalbių kontroliuojamų žodynų rengimo praktikos Lietuvoje analizė parodė, kad paprastai pasirenkamas pirmasis daugiakalbių tezaurų rengimo metodas, kai tam tikras tezauras, šiuo atveju HASSET, verčiamas į vieną ar kelias kitas kalbas. Šio metodo pasirinkimas reiškia, kad lietuviškas variantas yra „antrinis“, nes rengėjams teko derintis prie Jungtinės Karalystės HSM sampratos ir sistemos bei anglų kalbos struktūros. Be to, parengto dvikalbio tezauro specifinius bruožus lėmė konkretūs HASSET tezauro adaptavimo lietuvių kalbai sprendimai.

ELSST tezauras buvo rengiamas ne tik kaip daugiakalbis, bet ir daugiakultūris žodynas, siekiant, kad teiktini terminai atspindėtų Europos, o ne nacionalinį kontekstą. Nors ELSST tezaurui rengti taip pat buvo pasirinktas „bazinio“ tezauro vertimo į kitas kalbas metodas, vertimo samprata buvo specialiai patikslinta, akcentuojant tinkamo sąvokos reikšmės perdavimo svarbą. Todėl ELSST rengimas įgavo tam tikrų daugiakalbių tezaurų rengimo metodo bruožų, kai vienu metu rengiamos kelios kalbinės versijos. Be abejonės šis būdas suteikia didesnių garantijų, kad visos kalbos yra traktuojamos vienodai, taip išvengiant vienos kalbos ir kultūros dominavimo.

ŠALTINIAI IR LITERATŪRA

AITCHISON, Jean; GILCHRIST, Alan; BAWDEN, Dawid (2000). T*hesaurus Construction and Use*: A Practical Manual. 4th ed. London: Aslib.

BALKAN, Lorna (2010). Guide for ELSST Translators. User Guide for Access Level 4 [interaktyvus]. UKDA. 33 p. [žiūrėta 2010 m. gruodžio 15 d.]. Prieiga per internetą: <http://elsst.esds.ac.uk/UserGuide.aspx>.

BROUGHTON, Vanda (2006). *Essentials Thesaurus Construction.* London: Facet Pub., 2006. 289 p. ISBN 9781–85604–0/1–85604–565–X

Council of European Social Science Data Archives [interaktyvus] [žiūrėta 2011 m. sausio 18 d.]. Prieiga per internetą: <http://www.nsd.uib.no/cessda/>.

Data Documentation Initiative [interaktyvus] [žiūrėta 2011 m. sausio 18 d.]. Prieiga per internetą: <http://www.ddialliance.org>.

DOERR, Martin (2001). Semantic problems of thesaurus mapping [interaktyvus]. *Journal of Digital Informatikon,* Vol. 1, Issue 8 [žiūrėta 2011 m. sausio 28 d.]. Prieiga per internetą: <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/>.

European Language Social Science Thesaurus: issues in designing multilingual tool for social science research /Balkan, Lorna et al. [interaktyvus] [žiūrėta 2010 m. spalio 25 d.]. Prieiga per internetą: <http://crlao.ehess.fr/docannexe.php?id=1139>.

EUROVOC žodynas [interaktyvus] [žiūrėta 2010 m. gruodžio 20 d.]. Prieiga per internetą: <http://www3.lrs.lt/pls/inter/www_viewer.ViewTheme?p_int_tv_id=1637&p_kalb_id=1>.

ICPSR Subject Thesaurus. Inter university consortium for political and social studies [interaktyvus] [žiūrėta 2011 m. sausio 20 d.]. Prieiga per internetą: <http://www.icpsr.umich.edu/thesaurus/index.html>.

ISO 5964:1985 (1985) Documentation – Guidelines for the establishment and development of multilingual Ref. thesauri, International Organization for Standardization, No. ISO5964-1985.

ISO 2788-1986 (1986) Documentation – Guidelines for the establishment and development of monolingual thesauri, International Organization for Standardization, Ref. No. ISO 2788-1986.

JÄÄSKELÄINEN, Taina (2006). ELSST – Meeting the Challenge of a Multilingual Thesaurus [interaktyvus]. *FSD Bulletin* Issue 19 (1/2006) [žiūrėta 2010 m. lapkričio 12 d.]. Prieiga per internetą: <http://www.fsd.uta.fi/tietoarkistolehti/english/19/elsst.html>.

JÄÄSKELÄINEN, Taina; FORSMAN, Maria (2003). Finnish version of the Multilingual Thesaurus ELSST [interaktyvus] [žiūrėta 2011 m. sausio 12 d.]. Prieiga per internetą: <http://www.fsd.uta.fi/tietoarkistolehti/english/12/elsst.html>.

JORNA, Kerstin; DAVIES Sylvie (2001). Multilingual thesauri for modern world – no ideal solutions? *Journal of Documentation*, Vol. 57, No. 2, p. 284–295.

HASSET (Humanities and Social Science Electronic thesaurus) [interaktyvus] [žiūrėta 2011 m. sausio 28 d.]. Prieiga per internetą: <http://www.data-archive.ac.uk/find/hasset-thesaurus>.

HUDON, Michele (1997) Multilingual thesaurus construction-integrating the views of different cultures in one one gateway to knowledge and concepts. *Information Services & Use,* Vol. 17, Issue 2/3, p. 111–124.

KERÄNEN*,* Susanna (2002). Content Management – Concept and Indexing Term Equivalence in a Multilingual Thesaurus. Informing Science InSITE – "Where Parallels Intersect" June 2002 [interaktyvus] [žiūrėta 2011 m. sausio 15 d]. Prieiga per internetą: <http://informingscience.org/proceedings/IS2002-Proceedings/papers/keran122conte.pdf>.

Multilingual Access to European Cultural heritage: multilingual websites and thesauri (2006) [interaktyvus]. MINERVA content plus. 90 p. [žiūrėta 2010 m. spalio 17 d.]. Prieiga per internetą: <http://www.ifap.ru/library/book130.pdf>.

McCULLOCH, Emma (2005). Thesauri: practical guidance for construction. *Library Review,* Vol. 54, No. 7, 2005, p. 403–409.

McCULLOCH, Emma; SHIRI, Ali; NICHOLSON, Dennis (2005). Challenges and issues in terminology mapping: a digital library perspective. *The Electronic Library,* Vol. 23, No. 6, 2005, p. 671–677.

MENARD, Elaine (2010). A comparison of two indexing vocabularies. *ASLIB Proceedings,* Vol. 62, No. 4/5, p. 428–437.

MLSTEAD, Jessica; FELDMAN, Susan (1999). Metadata: Cataloging by Any Other Name … [interaktyvus]. *ONLINE*, Vol. 23, No. 1, p. 24–26, 28–31 [žiūrėta 2011 m. sausio 17 d.]. Prieiga per internetą <http://www.onlineinc.com/onlinemag/metadata>.

NIELSEN, Marianne (2001). A framework for work task based thesaurus design. *Journal of Documentation*, Vol. 57, No. 6, p. 774–797.

NYKYRY, Susanna (2010). Equivalence and Translation Strategies in Multilingual Thesaurus construction. [interaktyvus]. Abo, p. 433 [žiūrėta 2010 m. lapkričio 26 d.]. Prieiga per internetą: <https://oa.doria.fi/bitstream/handle/10024/59432/nykyri_susanna.pdf?sequence=1>.

SIHVONEN, Anne; VAKKARI, Pertti (2004). Subject knowledge improves interactive query expansion assisted by a thesaurus. *Journal of Documentation*, Vol. 60, No. 6, p. 673–690.

SHIRI, Ali; REVIE, Crawford; CHOWDHURY, Gobinda (2002). Thesaurus-assisted search term selection and query expansion: a review of user-centred studies. *Knowledge Organization*, Vol. 29, No. 1, p. 1–19.

SHIRI, Ali; REVIE, Crawford (2000). Thesauri on the Web: current developments and trends. *Online Information Review,* 2000, Vol. 24, No 4, p. 273–280.

TUDHOPE Douglas; BINDING Ceri; BLOCKS, Dorothee; CUNLIFFE, Daniel (2006). Query expansion via conceptual distance in thesaurus indexed collections. *Journal of Documentation,* Vol. 62, No. 4, p. 509–533.

The Thesaurus Review, Renaissance and Revision (2004). Eds. Roe, Sandra; Thomas Alan, Birghamton. N.Y, Haworth Infor. Pr. 209 p. ISBN 07890–1979–5.

VARNIENĖ, Regina ir kt. (2008) (Kuprienė, Jūratė; Prokopčik, Marija; Karvelytė, Vilma). Empirinių duomenų archyvavimo standartų ir dokumentavimo, duomenų aprašymo formato ir tezauro studija. Projektas „Empirinių duomenų ir informacijos HSM tyrimams kaupimas ir valdymas: Lietuvos HSM duomenų archyvas (LiDA)“. Kaunas. 83 p.

CONSTRUCTION AND USE OF MULTILINGUAL THESAURI: THE CASE OF THE LiDA LITHUANIAN DATA ARCHIVE

Marija Prokopčik

S u m m a r y

The article deals with the construction and use of multilingual linguistic tools – thesauri – in the electronic environment. The increasing diversity of languages used in the creation of intellectual product, growing amounts of people using these products and the variety of their cultural backgrounds turn multilingual thesauri into a specific multilingual and multicultural information retrieval.

The article aims to present the experience in constructing multilingual humanities and social science thesauri in Lithuania. The first one described is the Lithuanian–English thesaurus used in the Lithuanian data archive (LiDA). The main issue was to choose a proper method for creating the multilingual thesauri and then to analyse the thesauri existing in the field before selecting the most suitable one, namely the HASSET, to be used as a source dictionary. The article depicts the specific features of the constructed thesaurus, its use and maintenance.

The design and construction of the Lithuanian version of the multilingual and multicultural ELSST thesaurus is based on an analysis of the specific features of this thesaurus, the experience of its constructors and the general principles applied, such as cultural neutrality, equal treatment of all language versions, conceptual mapping, and adapting the terms for use in another language version.