



članek

VEČJEZIČNOST V KNJIŽNIČNIH INFORMACIJSKIH SISTEMIH

Petra BridgesInstitut informacijskih znanosti
MariborKontaktni naslov:
petra.bridges@izum.si**Izvleček**

Jezik je pomemben dejavnik gospodarskega razvoja, prevodi in lokalizacija pa so nujni za napredek in ekonomski razvoj. V Evropski uniji je jezikovno področje ključno, angleščina pa ni več dovolj za izgradnjo jezikovne politike, zato bi radi dosegli stanje, da bodo Evropejci poleg maternega jezika govorili še vsaj dva tuja jezika. Tudi v znanosti je zaznati vse več objav v maternem jeziku raziskovalcev in objava v angleščini ni več stvar prestiža. Jeziki naglo izumirajo, po ocenah naj jih okoli 3000 ne bi preživel tega stoletja, izguba jezika pa pomeni tudi izgubo identitete. Na spletu je aktivnih le 10 % jezikov, delež angleščine je le še 25 % in tako izgublja status lingve franke. Tudi digitalna ločnica vpliva na uporabo jezika na spletu; zaradi nje nekateri jeziki, ki imajo sicer veliko govorcev, na spletu sploh niso zastopani. Evropska unija priznava vse jezike držav članic kot uradne jezike EU in čeprav pospešeno razvija jezikovne tehnologije za avtomatizacijo prevajalskega procesa, te za zdaj še ne zadovoljujejo kriterijev kakovosti. Tudi digitalne knjižnice podpirajo večjezičnost, predvsem metapodatki igrajo pri tem pomembno vlogo. Standardizacija bibliografskega zapisa omogoča branje podatkov tudi v primeru, ko jezika zapisa ne obvladamo popolnoma. Večjezičnost je podprta tudi v sistemu COBISS, uporabniška oprema in dokumentacija sta dostopni v jezikih regije COBISS.Net, kar je pomemben prispevek k rebabilizaciji regije in ne nazadnje tudi sveta.

Ključne besede

večjezičnost, lingua franca, Evropska unija, umiranje jezikov, materinščina, svetovni splet, digitalna ločnica, metapodatki, COBISS, rebabilizacija

Abstract

Language is an important factor of economic development; translation and localisation are necessary for the overall progress and economic development. In the European Union, language is key and English is no longer sufficient for the development of a common language policy; they are now striving to achieve a situation where every European would speak at least two foreign languages. In science, the number of publications in the researchers' mother tongues has increased and publishing in English is no longer a sign of prestige. Languages are becoming extinct at an alarming rate; it is estimated that about 3000 will not survive this century. The loss of language also means the loss of identity to some extent. On the Internet, only about 10 % of world languages are active, and the share of English is only 25 % suggesting that it is losing its long-standing status as a lingua franca. The digital divide also has an impact on the use of languages on the Internet; because of that many large languages are visible on the Internet. The EU accepts the languages of all its member states as official EU languages. Even though language technologies for translation automation are being developed constantly, to this day they do not satisfy the quality criteria without any added human input. Digital libraries also support multilingualism with metadata playing an important part. The standardisation of bibliographic records allows for reading of data even in case one does not have perfect knowledge of the record language. Multilingualism is also supported in the COBISS system as the software and user documentation are available in the languages of the COBISS.Net region, which is an important contribution towards the rebabilisation of the region, and subsequently the world.

Keywords

multilingualism, lingua franca, European Union, language extinction, mother tongue, Internet, digital divide, metadata, COBISS, rebabilisation

REBABILONIZACIJA SVETA

Angleščina ni lingva franka

Znameniti "ocvrti piščanci iz Kentuckyja" so na Kitajskem najprej pogoreli, ker so znan slogan, s katerim so jih ponujali, ki se glasi "Dobro za prste polizat!", prevedli v "Da si poješ prste!". Američani do drugih jezikov pač niso pretirano pozorni, ker verjamejo v filozofijo "English-only", kar pa se spreminja. Mnogi so že dojeli, da ne moreš biti nosilec globalizacije, če se sam nisi globaliziral.

Jezik je postal pomemben dejavnik gospodarskega razvoja, čeprav ga večina ekonomistov še vedno obravnava zgolj kot strošek. Generalni direktorat za prevajalstvo Evropske unije je za leto 2008 izračunal, da je bila "jezikovna industrija" težka 8,4 milijarde evrov in da bo njena vrednost do leta 2015 narasla na 20 milijard. V izvoznem gospodarstvu je jasno: če ni prevoda, ni izdelka (Topping, 2000/111). Pri informacijski in komunikacijski tehnologiji govorimo o lokalizaciji, ki pomeni, da morajo biti prevedene vse verbalne vsebine in prilagojeni vsi neverbalni elementi, pri tem pa ne gre za izražanje vljudnosti, ampak za čisti ekonomski interes.

V študiji Univerze Stanford (Steele, 2009) so glede Kalifornije kot ene najbolj razvitih zveznih držav sveta ugotovili, da svoj uspeh v veliki meri dolguje svoji večjezičnosti in večkulturnosti. Na tem snujejo tudi svojo prihodnost, povezano z napredkom azijskih in južnoameriških držav (BRIC), kalifornijsko populacijo pa v večini sestavljajo prav ljudje teh kultur. Zato bodo dodatno vlagali v razvoj njihove kulture in jezikov.

Za EU je jezikovno področje eno od ključnih, zato ima Evropska komisija tudi komisarja za večjezičnost. V Komisiji so že pred leti ugotovili, da svoje jezikovne politike ne smejo graditi na angleščini, saj že kaže znamenja "omejenega roka trajanja". Nemški izvozniki so ugotovili, da z angleščino dobro shajajo v tridesetih državah, ampak obdelati jih morajo 80 ali 100, zato so začeli enakopravno upoštevati portugalščino, španščino, kitajščino, ruščino itd. EU vlaga v učenje tujih jezikov in namerava v doglednem času in prej kot drugi doseči stanje, ko bodo Evropejci poleg materinščine obvladali vsaj še dva druga jezika. To je tudi za Evropo zgodovinska prelomnica, saj je za t. i. zahodno civilizacijo veljalo načelo enojezičnosti: grščina v starem, latinščina v srednjem veku in nato kup agresivnih kandidatov za prevlado v novem veku. Danes pa se Evropa vneto zavzema za večjezičnost kot sestavino "kulturne inteligence" ali "kulturnega kvocienta" (angl. *Cultural Quotient – CQ*). "Jezik Evrope je prevod." pravi Umberto Eco in lingva franka je večjezičnost (Oustinoff, 2012, 55).

Lingua franca ima zanimivo, a manj znano zgodovino. Jezik je nastal v mediteranskih pristaniščih kot posebna oblika latovščine, neke vrste "pidžin" jezik, v katerem pa so se mešale besede iz italijanščine, francoščine, španščine, pa tudi grščine in arabščine. Dejansko je bil to nikogaršnji materni jezik in je služil čisto praktičnim namenom osnovne komunikacije. Tudi latinščina kot lingva franka se je bistveno razlikovala od angleščine in drugih vsiljenih jezikov po tem, da je bila nikogaršnja materinščina. Izpopolnjevali so jo kot orodje za prenos vsebin iz enega jezika preko latinščine v drug jezik – šlo je za trans-latiniranje, trans-latino.

Zdi se, da v znanosti prevladujejo tisti, ki jim angleščina povsem zadošča in imajo zanjo same pohvalne besede, da je enostavna, ekonomična, demokratična, nevtralna ipd. Von Humboldt je sicer že zdavnaj svaril, da nobenega jezika ni mogoče brez ostanka in škode pretvoriti v drugega, vendar mnoge znanosti tega problema ne zaznajo in sprejemajo angleščino kot polnokrvni jezik znanosti. Vendar pa se je pojavil drug problem, da je namreč vse več vrhunskih znanstvenih dognanj objavljenih v "neznatstvenih" jezikih: Kitajci res objavijo 110.000 znanstvenih člankov v angleščini, toda 470.000 jih objavijo samo v kitajščini (podatki za leto 2008). Se lahko kdo zanese, da prav v slednjih ni najobetavnejših rezultatov? Vse več člankov v domačih jezikih objavljajo tudi znanstveniki iz drugih azijskih in južnoameriških držav, pa tudi iz Evrope. Treba jih je obdelati, hraniti in brati. Zato pa je potrebna večjezičnost, ki je ni mogoče več jemati kot božjo kazen za prepotentne graditelje Babilonskega stolpa, ampak kot temeljno značilnost te naše "svetovne vasi".

Umiranje jezikov

National Virtual Translation Center je leta 2006 predstavil naslednje podatke o jezikih po kontinentih:

Področje	Število jezikov	% vseh jezikov	Število govorcev	% vseh govorcev
Afrika	2.092	30,3 %	675.887.158	11,8 %
Amerika	1.002	14,5 %	47.559.381	0,8 %
Azija	2.269	32,8 %	3.489.897.147	61,0 %
Evropa	239	3,5 %	1.504.393.183	26,3 %
Oceanija	1.310	19,0 %	6.124.341	0,1 %
VSI	6.912	100 %	5.723.861.210	100 %

Tabela 1: Razdelitev jezikov po kontinentih (Vir: Ching-chih Chen, 2007).

Tabela potrebuje kratko pojasnilo, ker je na prvi pogled neverjetna: število govorcev se močno razlikuje od števila prebivalcev, ki jih je bilo leta 2006 v Afriki 924 milijonov, v Severni Ameriki 332 milijonov, v Srednji

in Južni Ameriki 566 milijonov, v Aziji 3.968 milijonov, v Oceaniji 34 milijonov in v Evropi 732 milijonov (PRB, 2006). Vsi, ki govorijo izvorno evropske jezike (angleščino, španščino, portugalsščino, italijanščino itd.), so vsi med evropske govorce. Najbolj razširjeni jeziki leta 2006 so bili:

- mandarinščina (kitajščina), ki jo je kot svoj prvi ali drugi jezik govorilo 1.051 milijonov ljudi;
- angleščina, ki jo je kot prvi ali drugi jezik govorilo 508 milijonov ljudi;
- hindi/urdo, ki jo je kot prvi ali drugi jezik govorilo 466 milijonov ljudi;
- arabščina (vse variante), ki jo je govorilo 452 milijonov ljudi;
- španščina, ki jo je kot prvi ali drugi jezik govorilo 382 milijonov ljudi.

V 74 jezikih se izraža kar 94 % zemljanov, 70 % jezikov pa govorijo v 20 najrevnejših državah sveta. Če dodamo, da 90 % svetovnega bruto družbenega proizvoda (BDP) proizvedejo govorci 14 jezikov (angleščine, kitajščine, japonščine, nemščine, španščine, francoščine, italijanščine, ruščine, portugalsščine, arabščine, nizozemščine, korejščine, turščine in poljščine), nas lahko resno zaskrbi, kako ohraniti jezikovno raznolikost (GG inc., 2010).

Dramatično je, da jeziki naglo izumirajo in da je po pesimističnih ocenah vsaka dva tedna en jezik manj, po optimističnih ocenah pa jih izgine zgolj deset na leto. Pri tem je treba imeti pred očmi dejstvo, da 50 % jezikov redno uporablja manj kot 10.000 govorcev (Crystal, 2002). Jeziki, ki se uporabljajo le ustno in niso zavarovani s pisavo in slovnico, so v popolni negotovosti, je pa takih več kot 90 %. Unescov Atlas ogroženih svetovnih jezikov izrecno navaja okoli 3000 jezikov, ki ne bodo preživeli tega stoletja (Moseley, 2010). Umrlo je na tisoče jezikov in le z največjimi naporji so jih nekaj priklicali nazaj v življenje, recimo hebrejščino, katalonščino, baskovščino, francoščino v Quebecu itd.

Izguba jezika pomeni izgubo bistvenega elementa identitete, kar sproža osebno stisko in družbeno nestabilnost. Z veliko gotovostjo je mogoče napovedati, da bo prav to ena od razvojnih značilnosti človeštva v prihajajočih desetletjih. Unesco skrbno razvija metodologijo ocenjevanja vitalnosti jezikov, pri čemer so posebej pozorni na zmožnost jezika, da pokrije vse potrebe vsakdanje komunikacije. Zelo pomembna je jezikovna praksa otrok in mladih ljudi (Gibson, 2012, 78).

Materinščina je jezik, ki ga otrok bere z ustnic svoje matere, kar je seveda metafora, saj je razmerje med jezikom očeta in jezikom matere veliko bolj zapleteno.

Realnost sveta postaja diglotizem (dvojezičnost) in celo poliglotizem (večjezičnost). Če se spomnimo večjezične Jugoslavije, je bila materinščina pogosto drugi ali celo tretji jezik. Odločilno je, kakšno podporo uživa materinščina v izobraževanju, kodifikaciji jezika, pa tudi v IKT in drugje. Slovenija je z 91,1 % prebivalstva, ki uporablja slovenščino kot prvi jezik, v skupini evropskih držav jezikovno najbolj homogena (Krek, 2011, 9). Slovenščina je že skoraj stoletje varovana kot uradni jezik in kot obvezni jezik izobraževanja, razvite pa so tudi vse institucije, potrebne za ohranjanje in posodabljanje jezika. Ne glede na to so Slovenci skozi zgodovino prihajali v tesnejši stik najprej z nemščino, nato s srbohrvaščino in danes z angleščino, kar ni brez posledic za slovenščino. Velja ocena, da lahko 92 % prebivalcev v starosti od 25 do 64 let komunicira vsaj še v enem jeziku poleg materinščine (SURs, 2007). Potemtakem je večjezičnost realnost tudi za Slovenijo, saj ne more računati na velik vpliv slovenščine v mednarodnem prostoru.

VEČJEZIČNOST IN SVETOVNI SPLET

Jezik med spletom in muzejem

Unescova akcija Initiative B@bel izhaja iz ocene, da je na spletu le 10 % jezikov aktivnih. Vendar pa ni točna ocena, da je splet absolutna domena angleščine, saj se njen delež hitro zmanjšuje in danes predstavlja le še 25 % (Prado, 2012, 42). Marca 2011 je Google upošteval že 50 jezikov, kar pa je odvisno od prizadevnosti samih govorcev: islandščina z vsega 240.000 govorcev je prisotna od samega začetka, nekateri jeziki s 100 in več milijoni govorcev (bengalsščina, javanščina, tamilščina, haussščina, jaruba, itd.) pa niso zastopani. Wikipedia (v bazi ima že blizu 20 milijonov gesel) upošteva že 300 jezikov. Deleži se spreminjajo in vidno se veča prisotnost španščine, portugalsščine, ruščine, arabščine, še posebej pa kitajščine.

O prisotnosti nekega jezika na spletu ne odloča zgolj njegov kulturni potencial in število govorcev, ampak tudi tehnologija. IKT je nastala v območju angleškega jezika in latinske pisave in to prednost drugi težko nadoknadijo. Računalniški jeziki temeljijo na angleščini in računalniški strokovnjaki morajo to vzeti v zakup. Ta slika ni pretirano drugačna kot pri tisku, saj je le 30 jezikov, v katerih natisnejo več kot 1000 novih knjižnih naslovov letno. Dve tretjini svetovne literature izideta samo v šestih jezikih. V angleščini natisnejo 28 % knjig, kar pomeni, da je na papirju močnejša kot na računalniških zaslonih (Leclerc, 2011).

Specifična je tudi miselnost znanstvenikov, ki gledajo na uporabo angleščine kot na stvar prestiža in jo štejejo med svoje pomembne veščine – objava v angleščini

naj bi sama po sebi veljala več od objave v domačem ali katerem drugem jeziku. Le en angleški govorec je dovolj, da cela skupina sogovorcev preide na angleščino. Vendar pa samozavest pri govorcih drugih jezikov raste in je takega prostovoljnega podrejanja vse manj, kar napoveduje izrazitejšo večjezičnost tudi v znanosti.

Na prisotnost jezika na spletu vpliva digitalna ločnica (angl. *digital divide*), kar običajno razumemo kot omejen dostop do IKT: svetovni splet uporablja 10 % Afričanov, 25 % Azijcev, 65 % Evropejcev, 80 % Američanov ... Vendar sam dostop še ni dovolj, tehnologija mora biti tudi ustrezno prilagojena, kar dobro prikazuje uporaba pisav na spletu: latinično pisavo uporablja 39 % svetovne populacije, ki porabi 72 % papirja za tisk in zaseda 84 % interneta; kitajsko pisavo uporablja 22 % svetovne populacije, ki porabi 23 % papirja in zaseda 13 % interneta; arabsko pisavo uporablja 9 % svetovne populacije, ki porabi 0,5 % papirja in zaseda 1,2 % interneta; cirilico uporablja 5 % svetovne populacije, porabi 1,1 % papirja in zaseda 1,6 % interneta; indijsko pisavo uporablja 22 % svetovne populacije, ki porabi 2,2 % tiskarskega papirja in zaseda 0,3 % interneta (Mikami, 2012).

Informacijska družba je z vidika preživetja jezikov kontroverzna: s svojo pasivno držo lahko pospeši njihovo umiranje, aktivni spoprijem z novimi mediji pa lahko predstavlja "drugo rojstvo jezika". Najslabša pot, ki pa jo ubirajo mnogi tradicionalisti, je pričakovanje, da bo šla "tehnološka karavana" slej kot prej mimo in se bo svet povrnil v dobre stare Gutenbergove čase. To vodi v "tehnološki lingvici", jezikomor, ki se mu lahko izognemo le s trubarjansko ustvarjalnostjo in vztrajnostjo, ki je pred pol tisočletja postavila slovenščino med knjižne jezike. Danes ji je treba na podoben način zagotoviti mesto med spletnimi jeziki (Pivec, 2008). Ali kot vidi grožnjo Éric Poncet: "Astrofizikalna analogija s črno luknjo ni nobeno pretiravanje: jeziki brez zadostne inercije ali mase za vztrajanje na njihovi trajektoriji bodo neizogibno absorbirani in manjša kot bo njihova masa, prej se jim bo to zgodilo." (Poncet, 2012, 72).

Bistvenega pomena je mednarodna standardizacija IKT in spleta. Internet je nadaljevanje pismenosti in vemo, na kakšne težave je naletela slovenščina s svojimi šumniki. Nabor znakov (angl. *character set*) je problem, ki se rešuje že desetletja; od leta 1969, ko so ZDA uveljavile ASCII (American Standard Code for Information Interchange), pa do leta 2010, ko naj bi UNICODE 6.0 končno zajel vse abecede sveta, kar predstavlja nabor 109.449 črk (UC, 2010). (UNICODE je v celoti usklajen s standardom ISO/IEC 10646.) Velike zasluge za to ima Unesco, ki je leta 2003 sprejel Priporočila glede uveljavljanja in uporabe večjezičnosti in splošne dostopnosti spleta (Unesco, 2003).

Strojno prevajanje

Evropska unija načelno priznava jezike svojih članic, kar pomeni 23 jezikov ali 506 jezikovnih kombinacij. Ko bo združena cela Evropa, bo na kupu 60 jezikov in skoraj 4.000 jezikovnih kombinacij. Samo v Evropskem parlamentu bi danes potrebovali 8.500 prevajalcev, saj se letno nabere okoli 7 milijonov strani gradiva. Tega brez pospešenega razvoja "jezikovnih tehnologij" ne bo mogoče obvladati.

V tem trenutku "jezikovne tehnologije" ne zadovoljujejo kriterijev kakovosti. V mislih imamo procesiranje zapisanega jezika: razumevanje besedila, generiranje besedila, avtomatsko povzemanje, ekstrakcijo terminologije, informacijsko preiskovanje, Q&A sisteme, strojno prevajanje itd. Pri govorjenem jeziku gre za prepoznavanje in razumevanje govora, pretvarjanje govora v pisavo, sintetiziranje govora, prepoznavanje govora, identifikacijo govorjenega jezika, prevajanje govora, tolmačenje v realnem času itd. EU vlaga veliko truda v pospeševanje razvoja na tem področju, čemur je namenjen projekt META-NET. Ocena kakovosti strojnega prevajanja med jeziki EU je razvidna iz naslednje tabele, pri čemer velja, da "živi" prevajalci dosegajo rezultat okoli 80,0 in da postane avtomatski prevod uporaben šele pri rezultatu nad 50,0.

	Target Language																					
	en	bg	de	es	da	el	et	fi	fr	hu	it	lv	mt	nl	pl	pt	ro	sk	sl	sv		
en	-	40,5	46,8	52,6	50,0	41,0	35,2	34,8	38,6	50,1	37,2	50,4	39,6	43,4	39,8	52,3	49,2	55,0	49,0	44,7	50,7	52,0
bg	61,3	-	38,7	39,4	39,6	34,5	46,9	25,5	26,7	42,4	22,0	43,5	29,3	29,1	25,9	44,9	35,1	45,9	36,8	34,1	34,1	39,9
de	53,6	26,3	-	35,4	43,1	32,8	47,1	26,7	29,5	39,4	27,6	42,7	27,6	30,3	19,8	50,2	30,2	44,1	30,7	29,4	31,4	41,2
es	58,4	32,0	42,6	-	43,6	34,6	48,9	30,7	30,5	41,6	27,4	44,3	34,5	35,8	26,3	46,5	39,2	45,7	36,5	43,6	41,3	42,9
da	57,6	28,7	44,1	35,7	-	34,3	47,5	27,8	31,6	41,3	24,2	43,8	29,7	32,9	21,1	48,5	34,3	45,4	33,9	33,0	36,2	47,2
el	59,5	32,4	43,1	37,7	44,5	-	54,0	26,5	29,0	48,3	23,7	49,6	29,0	32,6	23,8	48,9	34,2	52,5	37,2	33,1	36,3	43,3
et	60,0	31,1	42,7	37,5	44,4	39,4	-	25,4	28,5	51,3	24,0	51,7	26,8	30,5	24,6	48,8	33,9	57,3	38,1	31,7	33,9	43,7
fi	52,0	24,6	37,3	35,2	37,8	28,2	40,4	-	37,7	33,4	30,9	37,0	35,0	36,9	20,5	41,3	32,0	37,8	28,0	30,6	32,9	37,3
fr	49,3	23,2	36,0	32,0	37,9	27,2	39,7	34,9	-	29,5	27,2	36,6	30,5	32,5	19,4	40,6	28,8	37,5	26,5	27,3	28,2	37,6
hu	64,0	34,5	45,1	39,5	47,4	42,8	60,9	26,7	30,0	-	25,5	56,1	28,3	31,9	25,3	51,6	35,7	61,0	43,8	33,1	35,6	45,8
it	48,0	24,7	34,3	30,0	33,0	25,5	34,1	29,6	29,4	30,7	-	33,5	29,6	31,9	18,1	36,1	29,8	34,2	25,7	25,6	28,2	30,5
lv	61,0	32,1	44,3	38,9	45,8	40,6	57,2	25,0	29,7	52,7	24,2	-	29,4	32,6	24,6	50,5	35,2	56,5	39,3	32,5	34,7	44,3
mt	51,8	27,6	33,9	37,0	36,8	26,5	41,0	34,2	32,0	34,4	28,5	36,8	-	40,1	22,2	38,1	31,6	31,6	29,3	31,8	35,3	35,3
nl	54,0	29,1	35,0	37,8	38,5	29,7	42,7	34,2	32,4	35,6	29,3	38,9	38,4	-	23,3	41,5	34,4	39,6	31,0	33,3	37,1	38,0
pl	72,1	32,2	37,2	37,9	38,9	33,7	48,7	26,9	25,8	42,4	22,4	43,7	30,2	33,2	-	44,0	37,1	45,9	38,9	35,8	40,0	41,6
pt	56,9	29,3	46,9	37,0	45,4	35,3	49,7	27,5	29,8	43,4	25,3	44,5	28,6	31,7	22,0	-	32,0	47,7	33,0	30,1	34,6	43,6
ro	60,8	31,5	40,2	44,2	42,1	34,2	46,2	29,2	29,0	40,0	24,5	43,2	33,2	35,6	27,9	44,8	-	44,1	38,2	38,2	39,8	42,1
sk	60,7	31,4	42,9	38,4	42,8	40,2	60,7	26,4	29,2	33,2	23,8	52,8	28,0	31,5	24,8	49,3	34,5	-	39,4	32,1	34,4	43,9
sl	60,8	33,1	38,5	37,8	40,3	35,6	50,4	24,6	26,2	46,5	25,0	44,8	28,4	29,9	28,7	43,0	35,8	48,5	-	31,5	35,1	39,4
sv	60,8	32,6	39,4	48,1	41,0	33,3	46,2	29,8	28,4	39,4	27,4	41,8	33,8	36,7	28,5	44,4	39,0	43,3	35,3	-	42,6	41,8
tr	61,0	33,1	37,9	43,5	42,6	34,0	47,0	31,1	28,8	38,2	25,7	42,3	34,6	37,3	30,0	45,9	38,2	44,1	35,8	38,9	-	42,7
uk	58,5	26,9	41,0	35,6	46,6	33,3	46,6	27,4	30,9	38,9	22,7	42,0	28,2	31,0	23,7	45,6	32,2	44,2	32,7	31,3	33,3	-

Tabela 2: Dosežki strojnega prevajanja med jeziki EU, merjeno v % BLEU (Algoritem BLEU ocenjuje kakovost besedila, ki je bilo strojno prevedeno iz enega naravnega jezika v drugega.) (Vir: Koehn, 2009)

Slovenščina v tej tabeli ne kaže pretirano dobrih rezultatov, sta pa Laboratorij za umetno inteligenco Instituta Jožef Stefan in podjetje Amebis zelo aktivna v projektu META-NET 2010–2013 (Multilingual Europe Technology Alliance, <http://www.meta-net.eu>) in si lahko obetamo napredek. Tudi pri tem je odločilna aktivna prisotnost jezika na spletu, saj je potreben čim večji besedilni korpus. V projektu se ukvarjajo s 30 jeziki in kar za 21 od njih kritično ugotavljajo, da imajo preslabotno podporo. Za islandščino, latvijščino,

litovščino in malteščino so mnenja, da so to jeziki v kritičnem stanju. Edini jezik z "odlično podporo" je angleščina. "Povprečno podporo" uživajo nizozemščina, francoščina, nemščina, italijanščina in španščina. Pri slovenščini, baskovščini, bolgarščini, katalonščini, grščini, madžarščini in poljščini ugotavljajo raziskovalci "delno podporo". Hans Uszkoreit, koordinator mreže META-NET pa zaključuje: "Rezultati študije so skrb vzbujajoči. Večina evropskih jezikov je digitalno slabo opremljenih, nekateri so popolnoma zapostavljeni. V tem smislu preživetje mnogih jezikov pravzaprav še ni zagotovljeno" (G.P., 2012). Očitno se tudi v Evropi pogloblja prepad med "velikimi" in "malimi" jeziki, saj se govorniki slednjih ne zavedajo, kako pomembna je jezikovna koda za splošno uspešnost družbe in se jim zdijo izdatki za ta namen odvečni. Še pomembnejša od denarja pa je dobra organiziranost, saj je kar značilno, da manjša ko je država, bolj razdrobljena in neučinkovita je skrb za vstop lastnega jezika v digitalno orbito.

Večjezičnost naj bi bila božja kazen za babilonske domišljavce, toda Evropski parlament je leta 2009 napovedal "enotni evropski informacijski prostor", ki ne bo "talilni lonec" različnih jezikov, ampak na vse strani prevedljivo pogovarjanje. Če smo v uvodu navedli Ecovo predvidevanje, da je "jezik Evrope prevod", moramo k temu dodati še "s pomočjo tehnologije".

KNJIŽNICE IN VEČJEZIČNOST

Ponovna pozidava slavne Aleksandrijske knjižnice se je zgodila v času, ko ni bila več potrebna – vsi viri znanja tega sveta so na spletu. Vloga "Alexandrine" igra danes globalna knjižnična mreža. Njeno "tehnološko okolje" ne predstavlja nobenega problema več, drugo pa je vprašanje, ali razumemo jezike, v katerih nam ponuja svoje vire? Brez večjezičnosti je zahtevno in drago povezovanje knjižnic vprašljivo.

Prvič so se s vprašanjem univerzalne izmenljivosti bibliografskih podatkov spoprijeli konec 19. stoletja, ko sta Belgijca Paul Otlet in Henri La Fontaine ustanovila Bibliografski inštitut (IIB), kar je bila podlaga za kasnejšo Mednarodno federacijo za dokumentacijo (FID). Začeli so graditi univerzalno zbirko metapodatkov o knjižničnih virih, ki je temeljila na univerzalni decimalni klasifikaciji (UDK), uporabni v vseh jezikih. Ob vstopu v novo stoletje je bila klasifikacija že tako razčlenjena, da je obsegala preko 2.000 strani. Kljub okorni tehnični podpori je Répertoire Bibliographique Universel že do prve svetovne vojne narasel na 11 milijonov zapisov. Ideja je bila, da bi zagotovili svetovni katalog, v katerem bo mogoče brez jezikovnih ovir preiskovati vse vire znanja. Ambicija je živela natančno 100 let, na koncu pa se je izkazala za preveč zapleteno, predvsem pa se

je pojavila računalniška tehnologija, ki je vse dotedanje predstave in rešitve postavila na glavo. UDK sicer še obstaja, FID kot njegova matična organizacija pa je prenehal delovati (Rayward, 1994).

Digitalne knjižnice podpirajo staro idejo na popolnoma nov način. Zmotno je, če v tem vidimo tradicionalno knjižničarstvo, ki stari pisalni stroj zgolj zamenjuje z računalnikom. In zmotno je, če v tem vidimo navaden računalniški spomin mimo bibliografskih standardov. Dokumenti, ki tvorijo knjižnično zbirko, so izbrani tako, da so v zbirki razumno zastopane vse bistvene ideje, teorije, perspektive in jeziki in bodo s tega vidika zdržali presojo časa in kritičnega uma. Bolj ko zbirka postaja digitalna, pomembnejša je selektivnost, saj "enkratnost" dokumentov ni več pomembna, ker jih je možno brezmejno kopirati, odločilna pa je "enkratnost" zbirke. Treba je natančno vedeti, za koga jo knjižnica organizira.

Knjižnica vire znanja odbira, hrani, organizira, popisuje in daje v uporabo. Vsa ta njena dejavnost se iz obdobja v obdobje spreminja, kot se razvijajo tehnologije, ki jo podpirajo. Kako prepleteni in kompleksni so ti procesi, je lepo razvidno na ravni bibliografskih metapodatkov, kar opisuje Hervé Le Crosnier: "Metapodatki zajemajo vse informacije, ki jih lahko pridobimo o dokumentu. Prvenstveno gre za deskriptivno katalogizacijo informacij: reference o avtorju, ilustratorju, uvodničarju; indikacije o izidu (datum objave, zbirka, reference o založniku, datum skeniranja itd.), lokacijski podatki o delu (število strani, značilnosti vezave, zgodovina ex libris, metoda skeniranja, oblika datoteke, trajanje zvoka ali avdio-vizualnega posnetka) in končno reference o povezanosti knjige z drugimi deli znotraj iste "družine" (vključno z indikacijami o izvorni izdaji, če gre za izvirnik s prevodom, za delo v več zvezkih ipd.). Deskriptivna katalogizacija opremlja gradivo z izdajateljskim kontekstom, bodisi da je najprej natisnjeno (CD, DVD) ali objavljeno neposredno online. Ta vidik konteksta je pri digitalnih dokumentih še posebej pomemben. Pogosto namreč lahkotnost uporabe samega besedila za preiskovanje dokumenta povzroči, da pozabimo dani dokument postaviti v širši kontekst (datum in pogoji objave, tipologija dokumenta, žanr dokumenta ipd.). Digitalna produkcija nam z enkratno možnostjo povezovanja informacij ("linkanja") omogoča, da gremo še veliko dlje, kot je običajno v "opredmetenih" knjižnicah. Povezovanje z avtorjevo biografijo, fotografijami, reprodukcijami naslovnih vseh jezikovnih verzij iste knjige in objavljenimi kritikami širi perspektive, umestitev in kontekst celotne objave" (Le Crosnier, 2012, 193).

Metapodatki so element večjezičnosti v knjižnici, saj so v številnih elementih posnetek dela, ki ga prikazujejo, in

so lahko v drugem jeziku. Standardizacija bibliografskega zapisa omogoča, da ga lahko "beremo" tudi z nepopolnim znanjem jezika, v katerem je zapisan. S tem se polje preiskovanja virov znanja močno razširi. Vendar pa je bila potrebna večja fleksibilnost opisa, zato je nastal model za opis virov RDF (Resource Description Framework), ki lahko zajame vse jezike in je resnični prag večjezičnosti v knjižničarstvu. Indeksiranje, ki temelji zgolj na programski opremi – OCR (Optical Character Recognition) – ki ga generira iskalnik, ko preiskuje besedila, se ne more kosati s profesionalno katalogizacijo, ki zna dodajati informacije iz konteksta. Skoraj gotovo bo šel razvoj v smeri sinergijske kombinacije avtomatskega indeksiranja in katalogizacije (He, 2006). Odkar obstaja model FRBR (Functional Requirements for Bibliographic Records) pa je mogoče dodajati različne verzije dela, tudi prevode, kar še dodatno razširi večjezičnost knjižničarstva (Tillet, 2004).

Digitalna knjižnica je korak do globalne knjižnice, in šele ta bo pravi odgovor knjižničarstva na izziv, ki ga predstavlja Google. Njena temeljna prednost mora biti večjezično preiskovanje vseh knjižničnih zbirk. Jezik je torej ključna perspektiva razvoja knjižničarstva in se kaže na petih tehnoloških ravneh:

- pragmatična raven – kulturni in družbeni vidiki ter konvencije,
- semantična raven – konceptualizacija,
- besedilna raven – identifikacija jezika in indeksiranje,
- simbolna raven – problemi kodifikacije,
- fizična raven – izgradnja in vzdrževanje mreže.

Globalna knjižnica bo večjezična, večkulturna in večpredstavnostna (multimedijska). Čakajo nas resne razprave o razmerju med lokalno in globalno razsežnostjo knjižničarstva, kot Damoklejev meč pa bo nad temi razpravami visel Google, ki ob pretiranem obotavljanju knjižničarjev lahko simbolizira tudi njihov umik v zgodovino.

Glede na odločilno vlogo večjezičnosti v razvoju Googlea mora knjižničarstvo budno spremljati tehnologijo večjezičnih slovarjev, avtomatskega (strojnega) prevajanja, večjezičnih iskalnikov, procesov lokalizacije programske opreme, izpolnjevanja vmesnikov itd. Temu je namenjen OLAC – Open Language Archives Community, ki zbira in organizira znanje o jezikih (Simons/Bird, 2003).

American Library Association (ALA) je leta 2007 sprejela Smernice za razvoj in uveljavitev večjezičnih zbirk in servisov, pri katerih izhaja iz odgovornosti knjižnic, da imajo vsi člani skupnosti enako raven knjižničnih servisov ne glede na etnično, kulturno

ali jezikovno poreklo. To ne sme biti kakšna dodatna storitev, ampak integralni del osnovne dejavnosti knjižnice. Knjižnica mora take smernice vgraditi v oblikovanje svoje zbirke (kriteriji selekcije, formati, bibliografski dostop, fizični dostop, vzdrževanje zbirke), v načrtovanje (kulturna raznolikost, program dejavnosti, večjezični informacijski in referenčni servis) in upoštevati pri zasedbi delovnih mest, ko je treba pritegniti govorce različnih jezikov v skupnosti (ALA, 2007).

VEČJEZIČNOST V SISTEMU COBISS IN PODOBNIH SISTEMIH

Pivec in Šercar (2007) navajata v članku, da se v regiji, ki jo zajema sistem COBISS, govori 8 jezikov in vsaj 10 "manjšinskih jezikov", v knjižnicah pa je ohranjenih še nekaj "mrtvih jezikov", kot so latinščina, grščina, starocerkvena slovanščina. Raznolike so tudi pisave, saj se piše v več verzijah latinice in cirilice, za potrebe katalogizacije pa knjižničarji potrebujejo tudi grško in arabsko pisavo.

Države, ki jih trenutno pokriva sistem COBISS v sklopu projekta COBISS.Net, so: Bolgarija, Bosna in Hercegovina, Črna gora, Makedonija, Slovenija, Srbija, Kosovo in po novem tudi Albanija.

Jezikov pa je še precej več kot držav. Samo v Sloveniji je uradni jezik slovenščina, na področjih, kjer živita italijanska in madžarska narodna manjšina, pa še italijanščina in madžarščina. Letošnja izdaja romsko-slovenskega slovarja prča o tem, da tudi delež romsko govorečega prebivalstva ni zanemarljiv.

Za 87 % prebivalcev Slovenije je slovenščina materni jezik, za preostalih 13 % prebivalcev pa tuji jezik. Statistični podatki za leto 2011 pravijo tudi, da je 93 % odstotkov ljudi v starosti 18–69 let govorilo vsaj en tuji jezik. Iz grafičnega prikaza je razvidno, da gre pri tem v veliki večini za angleški jezik, sledi mu nemščina.

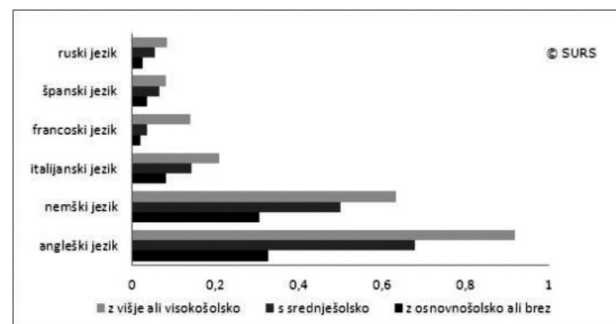


Tabela 3: Delež prebivalcev v starosti od 18 do 69 let po izobrazbi in znanju tujih jezikov (Vir: SURS, 2012)

V Bolgariji je uradni jezik bolgarščina, med manjšinske jezike pa sodita turški in romski jezik. V Srbiji je uradni jezik srbsčina, vendar govorijo tudi albanščino, romunščino, madžarščino, slovaščino, rusinščino, hrvaščino, ukrajinščino, bolgarščino, romščino, češčino, bosanščino in makedonščino. Albanščino govorijo tudi v Albaniji in na Kosovu. V Črni gori je uradni jezik črnogorski, vendar prebivalci govorijo tudi srbsko, albansko, bosansko in hrvaško. V Bosni in Hercegovini pa se poleg bosanskega govora še hrvaški in srbski jezik.

Uporaba jezikov v regiji je botrovala tudi odločitvi, da so programska oprema, vmesniki in uporabniška dokumentacija znotraj sistema COBISS v mreži COBISS.Net na voljo v več jezikih. Ekipa prevajalcev skrbi za prevod uporabniške opreme v šest jezikov (slovenski, angleški, srbski, makedonski, albanski in bolgarski). Priročniki za uporabo programske opreme se prevajajo v pet jezikov (slovenski, angleški, srbski, albanski in bolgarski). Spletne strani platforme COBISS.Net so prav tako dostopne v vseh uradnih jezikih držav, ki sodelujejo v tej mreži, vsebine pa so prilagojene vsaki državi posebej. Spletni katalog COBISS/OPAC je uporabnikom na voljo v slovenskem in angleškem jeziku, dostop do bolgarske in albanske bibliografske baze pa je možen tudi v bolgarskem in albanskem jeziku. Zanimiv je tudi porast jezikov, v katerih je objavljena spletna stran konference COBISS. Število jezikov je tako od leta 2003 (samo slovenski jezik) naraslo na šest (slovenski, angleški, srbski oz. črnogorski, makedonski, bolgarski in albanski) v letu 2011.

Če si na kratko ogledamo še nabor znakov COBISS, ki se uporabljajo pri vnosu bibliografskih podatkov in podatkov o zalogi v uporabniški opremi COBISS3, potem lahko vidimo, da za latinico vsebuje znake za 35 jezikov (albanski, beloruski, bolgarski, bosanski, češki, danski, dolnjelužiško srbski, esperanto, estonski, finski, francoski, furlanski, gornjeluziško srbski, hrvaški, islandski, italijanski, laponski, latvijski, litovski, lužiško srbski, madžarski, makedonski, nemški, norveški, poljski, portugalski, romunski, ruski, slovaški, slovenski, srbski, španski, švedski, turški in ukrajinski). Nabor znakov COBISS, prav tako opisan v priročniku za uporabnike programske opreme COBISS3/Katalogizacija, za cirilico vsebuje znake za sedem jezikov (beloruski, bolgarski, makedonski, rusinski, ruski, srbski in ukrajinski). Tukaj je treba omeniti, da so naštetni samo jeziki, ki v svojem naboru uporabljajo znake, ki niso del standardnega nabora (npr. angleškega jezika). (IZUM, 2012)

Največji spletni katalog z javnim dostopom na svetu, WorldCat, ki ga ustvarja in oskrbuje OCLC, se na svojih spletnih straneh ponaša s tem, da so podatki na voljo v več kot 470 jezikih in da vsebuje podatke iz 112 držav in teritorijev, njihova programska oprema za katalogizacijo

pa podpira arabsko, bengalsko, kitajsko in cirilično pisavo, poleg tega pa še devangari, grško, hebrejsko, japonsko, korejsko, tamilsko in tajsko, vsaka od teh pisav pa vključuje številne jezike. Njihovo orodje za iskanje WorldCat search pa ponuja iskanje v kar 30 jezikih.

Seveda se ponudba jezikov znotraj sistema COBISS s tem ne more primerjati, vendar pa COBISS kljub temu pomembno prispeva k ohranjanju večjezičnosti v zgodovinsko in kulturno zelo raznoliki regiji.

In če na koncu prosto povzamemo še misel Franka Smitha, da nas en jezik postavi na hodnik življenja, dva pa odpreta vsa vrata v njem, potem postane jasno, zakaj si bomo tudi v prihodnje prizadevali za čim večjo pokritost jezikov in čim boljše prevode tudi v sistemu COBISS.

Reference

- [1] ALA (2007). Guidelines for the Development and Promotion of Multilingual Collection and Services. Dostopno na: <http://www.ala.org/rusa/resources/guidelines/guidelmultilingual>.
- [2] Ching-chih Chen (2007). Delivery of Web-based multilingual digital collections and services to multicultural populations. IFLA Meeting in Durban, South Africa, August 18–24, 2007. Dostopno na: (<http://www.ifla.org/IV/ifla73/Programme2007.htm>).
- [3] Crystal, David (2002). *Language Death*. Cambridge: Cambridge University Press.
- [4] GG inc. (2010). Top Languages by GDP-2010. Dostopno na: <http://www.globalization-group.com/edge/2010/03/top-language-by-gdp/>.
- [5] Gibson, Maik (2012). Preserving the Heritage of Extinct or endangered Language. V: Vannini, L. and Le Crosniere, H. (ed.) *NET.LANG. Towards the multilingual cyberspace*. Caen: C&F Éditions.
- [6] G.P. (2012). Vsaj 21 evropskim jezikom grozi digitalno izumrtje. *DELO*, 28/9 2012.
- [7] He, Daqing, Oard, W.D., Plettenberg, L. (2006). Studying the Use of Interactive Multilingual Information Retrieval. SIGIR'06 Workshop, Seattle, August 11 2006. Washington: ACM.
- [8] IZUM (2012). COBISS3/Katalogizacija, V5.1-05: Priročnik za uporabnike.
- [9] Koehn, Ph. et al. (2009). Machine translation Systems for Europe. Machine Translation Summit XII. 2009, str. 65–72.
- [10] Krek, Simon (2011). Slovenski jezik v digitalni dobi. Dostopno na: <http://www.meta-net.eu/whitepapers/e-book/slovene.pdf>.
- [11] Leclerc, Jacques (2011). L'expansion des langues. V: L'aménagement linguistique dans le monde. Dostopno na: http://www.tlfq.ulaval.ca/axl/languages/2vital_expansion.htm.
- [12] Le Crosnier Hervé (2012). Digital Libraries. V: Vannini, L. and Le Crosniere, H. (ed.) *NET.LANG. Towards the multilingual cyberspace*. Caen: C&F Éditions.
- [13] Mikami, Yoshiki & Kodama, Shigeaki (2012). Measuring linguistic Diversity on the Web. V: Vannini, L. and Le Crosniere,

- H. (ed.) NET.LANG. Towards the multilingual cyberspace. Caen: C&F Éditions.
- [14] Moseley, Christopher (2010). Atlas of the World's Languages in Danger. Paris: Unesco Publishing.
- [15] Oustinoff, Michael (2012). English won't be the internet's lingua franca. V: Vannini, L. and Le Crosniere, H. (ed.) NET.LANG. Towards the multilingual cyberspace. Caen: C&F Éditions.
- [16] Pivec, Franci (2008). Jeziki na internetu. *Stati inu obstatu*, 7–8 (okt.). 2008, str. 113–120.
- [17] Pivec, Franci & Šercar, Trvrtko-Matija (2007). Vloga kulturnega konteksta pri nastajanju knjižničnih informacijskih sistemov v mreži COBISS.Net. *OZ*, 12 (4), 2007, str.206–21.
- [18] Poncet, Éric (2012). Technological Innovation and Language Preservation. V: Vannini, L. and Le Crosniere, H. (ed.) NET.LANG. Towards the multilingual cyberspace. Caen: C&F Éditions.
- [19] Prado, Daniel (2012). Language Presence in the Real World and Cyberspace. V: Vannini, L. and Le Crosniere, H. (ed.) NET.LANG. Towards the multilingual cyberspace. Caen: C&F Éditions.
- [20] PRB (2006). World Population Data Sheet. Washington: Population Reference Bureau.
- [21] Rayward, W. Boyd (1994). The International Federation for Information and Documentation (FID). V: Wayn, Wiegand and Davis (ur.) Encyclopedia of Library History. New York: Garland Press (290–294).
- [22] Simons, G. and Bird, S. (2003). Building an Open Language Archives Community on the OAI foundation. *Library Hi Tech*, 21, str. 210–218.
- [23] SURS – Statistični urad Republike Slovenije (2010). Izobraževanje odraslih po Anketi o izobraževanju odraslih 2007. Dostopno na: <http://www.stat.si/>.
- [24] SURS – Statistični urad Republike Slovenije (2012). Evropski dan jezikov 2012. Dostopno na: <http://www.stat.si/>.
- [25] Tillet, Barbara (2004). What is FRBR? A Conceptual Model for the Bibliographic Universe. Washington: Library of Congress.
- [26] Topping, Suzanne (2000). Shortening the Translation Cycle at Eastman Kodak. V: Sprung, R.C. (ed.) Translating into Success: Cutting-Edge Strategies for Going Multilingual in a Global Age. Amsterdam: John Benjamins.
- [27] UC – UNICODE Consortium (2010). The UNICODE Standard, Version 6.0.0. The UNICODE Consortium. Dostopno na: <http://www.unicode.org>.
- [28] Unesco (2003). Recommendation Concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace. Paris: Unesco.