

izvirni znanstveni članek  
prejeto: 2010-05-07

UDK 81'374.26:338.48

## IDENTIFIKACIJA TERMINOV ZA TURISTIČNI TERMINOLOŠKI SLOVAR

Vesna MIKOLIČ

Univerza na Primorskem, Fakulteta za humanistične študije Koper, SI-6000 Koper, Titov trg 5  
e-mail: vesna.mikolic@fhs.upr.si

Ana BEGUŠ

Univerza na Primorskem, Fakulteta za humanistične študije Koper, SI-6000 Koper, Titov trg 5  
e-mail: ana.begus@fhs.upr.si

### IZVLEČEK

*V prispevku je predstavljen postopek urejanja turističnega terminološkega slovarja s poudarkom na identifikaciji terminov, ki je potekal po korpusnem pristopu. Najprej je pozornost posvečena izboru in klasifikaciji besedil za korpus turističnih besedil TURK, ki danes obsega približno 30 milijonov besed. Na osnovi korpusa je bila avtomatsko generirana lista besed (monogramov) in besednih zvez (bigramov in trigramov), urejena po pogostosti od najvišje k najmanjši. Nadaljnji postopek luščenja turistične terminologije je bil izveden na osnovi predhodne razvrstitve turističnih besedil v korpusu glede na vrsto turizma in glede na družbeno področje, ki se vključuje v turizem. Prav družbeno področje, ki se v geselskem članku terminološkega slovarja izrazi kot kvalifikator posameznega termina, se je izkazalo za najbolj relevantno pri odločitvi za nabor turističnih terminov.*

**Ključne besede:** turizem, terminologija, terminološki slovar, terminološki kandidat, specializirani korpus, področna besedila, funkcijska zvrstnost, geselski članek

## IDENTIFICAZIONE DEI TERMINI PER IL VOCABOLARIO TERMINOLOGICO DEL TURISMO

### SINTESI

*Nell'articolo viene illustrata la procedura adottata nella compilazione di un vocabolario terminologico del turismo, con particolare riferimento al processo di identificazione dei termini avvenuta secondo l'approccio per corpus. Si è iniziato con la selezione e la classificazione dei testi per il corpus specialistico di turismo TURK che oggi si compone di 30 milioni di parole circa. In base al corpus è stata automaticamente composta la lista di parole (monogrammi) e di associazione di parole (bigrammi e trigrammi), organizzata per frequenza d'uso, dalla maggiore verso la minore. La successiva selezione della terminologia del turismo è stata eseguita in base alla precedente classificazione dei testi nel corpus a seconda della tipologia di turismo e del segmento sociale che interagisce con il turismo. Proprio il segmento sociale che nell'articolo descrittivo del lemma del vocabolario terminologico qualifica il singolo termine, si afferma quale aspetto più rilevante nella selezione dei termini del turismo.*

**Parole chiave:** turismo, terminologia, vocabolario terminologico, candidato termine, corpus specialistico, testi appartenenti a linguaggi settoriali, tipologia funzionale, articolo descrittivo dei lemmi

## POMEN TURISTIČNE TERMINOLOGIJE

Terminologija predstavlja pomemben del besedja vsakega modernega jezika, ki med drugim odraža dogajanja v določeni družbi, življenje neke skupnosti, suverenost narodnega jezika in naroda kot takega. Za sodobni čas je značilno izredno hitro širjenje informacij, nastajanje novih izdelkov, naprav, odkrivanje novih spoznanj. Vse to je potrebno poimenovati, zato imata terminologija in terminografija zelo pomembno vlogo. Tvorjenje strokovnega izrazja je v teoriji knjižnega jezika ena temeljnih nalog jezikovne skupnosti, ki svoj jezik uporablja na različnih sporazumevalnih področjih. Tako se z nastankom države Slovenije slovenski jezik uporablja na vseh sporazumevalnih področjih, tudi v vojski, carini, diplomaciji, kjer se je v nekdanji Jugoslaviji uporabljala srbohrvaščina. Slovenci se zavedamo pomembnosti terminologije – na to kaže veliko število terminoloških slovarjev, izdanih zlasti v 90-ih letih prejšnjega stoletja. Kljub temu pa se kažejo razlike med področji, ki so hkrati z razvojem skrbela za slovenska poimenovanja, npr. informacijska tehnologija in mobilna telefonija (Stramljič Breznik, 2003), in nekaterimi drugimi področji, ki pod vplivom globalizacijskega jezikovnega enotjenja in pod pritiskom državne politike ocenjevanja opuščajo razvijanje slovenske terminologije (Humar, 2009, 42).

Zaradi spoznanja, da turistična terminologija v slovenščini še nima nobenega enojezičnega ali večjezičnega slovarja ali leksikona, se je priprava terminološkega slovarja s področja turizma pokazala kot nujna. Slovenija namreč že od nekdaj velja za turistično deželo, z vstopom v Evropsko unijo pa je naša država kot posebej zanimiva turistična destinacija še pridobila pomen. Številni razgovori, ki smo jih med večletnim ukvarjanjem z oglaševanjem v turizmu opravili z nosilci in izvajalci turističnih dejavnosti, so pokazali, da so le-ti še kako zainteresirani za učinkovita orodja, ki bi pripomogla k učinkovitosti njihovih trženjskih procesov. Terminološki slovar na osnovi elektronskega večjezičnega jezikovnega korpusa kot statistično relevantne zbirke različnih turističnih besedil bo tako pomenil osnovo za delovanje celotne turistične stroke, in sicer tako neposrednih nosilcev turistične dejavnosti, tj. podjetij (hotel-skih, gostinskih, transportnih idr.), kulturnih centrov in društev, zvez kulturnih organizacij, kot tudi za posredne nosilce te dejavnosti, kot so različna poslovna združenja, turistična društva in zveze, lokalne skupnosti, gospodarske zbornice idr. Hkrati pa bo seveda pomembna osnova za vse vede, ki se s turizmom ukvarjajo kot s predmetom svojega preučevanja.

Terminološki slovarji so namreč atributi razvitih strok, omogočajo strokovno in jezikovno ustrezno komunikacijo, predlagajo slovensko izrazje za nove predmete in pojme hitro razvijajočih se ali novih strok, odražajo duha in kulturo naroda (če so narejeni na domačem gradivu!) ter odražajo politične tendence (Humar, 2004, 23). Hkrati večjezični terminološki slovarji omogočajo tudi mednarodno komuniciranje v novih družbenih in političnih razmerah.

Seveda se tu pojavi vprašanje zasnove in načina oblikovanja terminološkega slovarja. Logarjeva (2009, 321) navaja tri faze v nastajanju terminološkega slovarja, in sicer: opredelitev besedil, iz katerih bo pridobljena leksika; oblikovanje meril, na podlagi katerih bo leksika vključena v geslovnik; določitev podatkov, ki bodo vključeni v geselski članek. Pričujoči prispevek ima namen prikazati teoretske opredelitve in praktične rešitve, pomembne za vse tri faze nastajajočega turističnega terminološkega slovarja, s poudarkom na identifikaciji terminov.

Slovenski razlagalni turistični terminološki slovar z vzporednico v angleškem jeziku nastaja v okviru aplikativnega projekta "Turistični terminološki slovar" Znanstveno-raziskovalnega središča Univerze na Primorskem, financer: ARRS, nosilka: dr. Vesna Mikolič, 2008–2011. Kodificiranje slovenske turistične terminologije v okviru omenjenega projekta temelji na:

- evidentiranju strokovnih izrazov ali terminov,
- pomenski analizi terminov in ugotavljanju pomenskih odnosov med njimi (sopomenskost, protipomenskost, večpomenskost, nad- in podpomenskost),
- normiranju glede na knjižno normo in zahteve urejene terminologije.

Osnova za izdelavo slovarja je specializirani jezikovni korpus s področja turizma, ki zajema 30.000.000 besed in je nastal v okviru temeljnega raziskovalnega projekta "Večjezični korpus turističnih besedil – informacijski vir in analitična baza slovenske naravne in kulturne dediščine" istega izvajalca, financerja in iste nosilke v letih 2006–2008. Cilj projekta je bila izdelava elektronske zbirke turističnih besedil v slovenskem, italijanskem in angleškem jeziku, oblikovane na osnovi statistično relevantnih kriterijev in z nizom pomožnih jezikovnih orodij.

#### IZBOR IN KLASIFIKACIJA BESEDIL TURISTIČNEGA JEZIKOVNEGA KORPUSA KOT VIRA PODATKOV ZA TERMINOLOŠKI SLOVAR

Specializirani ali področni korpus<sup>1</sup> zajema besedila z določenega predmetno-pojmovnega področja, ki je lahko zelo razvejano in mnogoplastno. Takšen primer je

1 Za utemeljitev strokovnega izraza področni korpus glej Mikolič, 2009, 257–263.

tudi jezikovni korpus s področja turizma, pri gradnji katerega je potrebno upoštevati vsaj dve posebnosti turizma kot hitro razvijajočega se družbenega področja. Prvič, turizem kot tak zajema vse vidike družbe, saj kot množični pojav temelji na številnih gospodarskih in tudi negospodarskih dejavnostih. In drugič, turizem je področje, na katerem jeziki v globalizacijskih procesih sodobne evropske in svetovne družbe intenzivno prihajajo v stik z drugimi jeziki. Tako je turizem (poleg diplomacije, gospodarstva, znanosti, visokega šolstva – učiteljske in študentske izmenjave, vsakdanjega življenja na obmejnih območjih idr.) eno ključnih področij, kjer se tudi slovenski diskurz sooča s tujejezičnimi diskurzi. Zaradi teh specifičnosti turizma v terminologijo prihajajo tujejezične prvine ter izrazi iz drugih strok v stiku, ki jih je zato potrebno urediti skladno s sodobnimi načeli leksikografije in terminografije.

Tako se v primeru turistične področne sfere oblikujejo in uporabljajo različne besedilne zvrsti in vrste s specifičnimi jezikovnimi zvrstmi in sredstvi, pri čemer se določena jezikovna sredstva lahko istočasno pojavljajo v različnih besedilnih zvrsteh in vrstah. Npr. tipična sredstva področnega strokovnega jezika, npr. termine, lahko srečamo tako v znanstvenih, umetnostnih, pravnih in sporazumevalnih (strokovnih, publicističnih, oglaševalskih idr.) besedilih s tega področja. Ali obratno, npr. v oglaševalskih besedilih lahko srečamo sredstva, značilna za področje oglaševanja (npr. slogan), pa tudi določene stroke (npr. termini). Kaže se torej, da področje in v okviru tega oblikovana besedilna zvrstnost določata izbiro jezika, zato se sodobne usmeritve jezikoslovnega raziskovanja, ki postavljajo besedilo ali diskurz pred jezik, kažejo za utemeljene.

Pri gradnji večjezičnega korpusa turističnih besedil je bilo nujno ločevanje med besedili glede na turistična (pod)področja in vrste turizma ter glede na vplivajnsko vlogo in druge okoliščine sporazumevanja, v katerih se oblikujejo različne zvrsti in vrste besedil, kar je predstavljeno v nadaljevanju tega poglavja (povzeto po Mikolič, 2009, 257–263).

#### **Turistična besedila glede na (pod)področja in vrste turizma**

Turizem je izrazito sestavljena dejavnost, ki jo poleg turistične industrije kot turistične dejavnosti v ožjem pomenu besede sestavljajo tudi dejavnosti na skoraj vseh družbenih področjih, saj turizem kot družbeni pojav prežema najrazličnejše vidike družbe. Tako poleg turizma v ožjem pomenu besede vstopajo v okvir turistične dejavnosti tudi družbena področja in vede, kot so: ekonomija, pravo, geografija, zgodovina, umetnost, umetnostna zgodovina, arhitektura, urbanistično in regionalno planiranje, antropologija, sociologija, komunikologija, jezikoslovje, psihologija, etnologija, kemija, medicina, botanika, ekologija, agronomija, gozdarstvo,

veterinarstvo, lovstvo, ribištvo, kulinarika, šport, alpinizem, jamarstvo, javna uprava, zdravstveno varstvo in zdravstveno zavarovanje, javna varnost in policija, promet, logistika, zaposlovanje, vzgoja in izobraževanje, prostočasne dejavnosti, publicistika. Seznam je delno povzet po Goeldner in Brent Richie (2003, 23–25), dopolnjen je še z nekaterimi turističnimi področji, ki jih v okviru 20 pomenskih področij (*semantic fields*) navaja *Tezaver turizma in prostočasnih dejavnosti* Svetovne turistične organizacije (World Tourism Organization, 2001), ter nekaterimi področji, ki so se zdela še relevantna za turistično dejavnost ob pregledu področij dela vseh ministrstev slovenske vlade (spletne strani ministrstev).

Vsa ta družbena področja in vede se srečujejo v okviru različnih vrst turizma, ki izpostavljajo točno določeno turistično ponudbo. Ob pregledu obstoječih klasifikacij in delitev posameznih vrst turistične dejavnosti je bil za potrebe označevanja besedil v sklopu obravnavanega korpusa v veliki meri upoštevan *Tezaver turizma in prostočasnih dejavnosti* Svetovne turistične organizacije (2001), ki s svojim podrobnim vsebinskim razločevanjem turističnih dejavnosti (24 vrst turizma) jasno definira poglavitne zvrsti znotraj te družbene dejavnosti. Zaradi izrazite usmerjenosti posameznih vrst turizma na nekatera "nišna" področja znotraj turistične dejavnosti in hkrati smiselne vsebinske dopolnitve same klasifikacije za označevanje besedil smo izbrano zvrstno delitev Svetovne turistične organizacije dopolnili najprej še s šestimi dodatnimi vrstami turističnih aktivnosti ter splošno kategorijo "turizem", v katero so bila uvrščena besedila, ki jih bodisi zaradi pomanjkanja ključnih pojmov bodisi zaradi prepletanja več vrst turizma ni bilo mogoče umestiti v posamezno zvrst besedil. Tako se besedila glede na vrsto turizma v obravnavanem korpusu razvrščajo v naslednje zvrsti: religiozni turizem, kulturni turizem, poslovni turizem, pustolovski turizem, industrijski turizem, festivalski turizem, izobraževalni turizem, kmečki turizem, podeželski turizem, urbani turizem, naravni turizem, gorniški turizem, podzemski turizem, turizem bližnjih destinacij, zdraviliški turizem, turizem za osebe s posebnimi potrebami, mladinski turizem, družinski turizem, turizem za starejše, luksuzni turizem, avtodomni turizem, rečni turizem, obmorski turizem, morski turizem, zimski turizem, športni turizem, seksualni turizem, igralniški turizem, kulinarčni turizem.

V korpusu turističnih besedil smo besedila razvrščali samo glede na vrsto turizma in ne glede na družbeno področje, saj se različna področja vključujejo v različne vrste turizma. Določanje podpodročja turizma pa je postalo relevantno v procesu identifikacije terminoloških kandidatov (glej poglavje 3) pri pripravi terminološkega slovarja, saj je področje lahko pomembno kot kvalifikator posameznega termina.

### Turistična besedila glede na vplivansko vlogo in druge okoliščine sporazumevanja

Turistična besedila se z vidika osnovne vplivanske vloge razvrščajo med znanstvene, sporazumevalne in uradne besedilne zvrsti. Tako na osnovnem nivoju razlikovanja besedil glede na zvrstnost obravnavani korpus, enako kot referenčni korpus Fidaplu<sup>s</sup>, ločuje med 'umetnostnimi' in 'neumetnostnimi' besedili, pri čemer prvih korpus turističnih besedil ne vključuje. Besedila se dalje delijo na šest velikih skupin:

- pravna (besedila z izvršilno vlogo): to je edina skupina, ki ostaja le pri tej prvostopenjski oznaki; gre namreč za sorazmerno lahko prepoznavna besedila z visoko stopnjo konvencionaliziranosti oblike, kot so zakoni, pogodbe, odločbe, akti, sklepi ipd.;

- znanstvena (besedila s spoznavno vlogo): 'znanstvene monografije', 'znanstveni članki' ter 'magistrska in doktorska dela';

- strokovna: 'naravoslovna in tehnična' ter na 'humanistična in družboslovna', slednja pa še naprej na 'strokovne monografije', 'strokovne članke', 'poljudnoznanstvene monografije', 'poljudnoznanstvene članke', 'diplomske naloge', 'učbenike' in 'drugo';

- poslovna: 'korespondenca' (tu je mišljena običajna pošta) ter 'elektronska pošta';

- oglaševalska: 'oglasna sporočila', 'letake, prospekte, brošure, kataloge', 'prodajne oglase', 'oglasne na prostem', 'vodnike' in 'drugo';

- publicistična: 'novinarski članki' (npr. vest, poročilo, članek, komentar), 'strokovni novinarski članki', 'reportaže', 'članki v turističnih prilogah' in 'drugo'.

Vsa besedila sodijo med besedila s sporazumevalno vlogo.

Seveda se je zato pokazalo metabesedilno označevanje korpusa turističnih besedil za zelo zahtevno naloge; navkljub dokaj razdelanemu naboru oznak smo našli na veliko besedil, ki jih je bilo težko natančno in nedvoumno označiti. Najpomembnejši razlog za to sta gotovo kompleksnost in razvejanost turistične stroke oz. področja, ki prav zato potrebuje sistematično analizo tipologije turističnih besedil glede na različne kriterije. Poleg tega pa takšno metabesedilno razlikovanje in označevanje narekujejo tudi različni nameni gradnje korpusa, ki bo uporaben kot prevajalski vir, za raziskave turizma in različne jezikoslovne raziskave, predvsem pa kot osnova za izdelavo turističnega terminološkega slovarja.

### IZBOR TERMINOLOŠKIH KANDIDATOV

V procesu izbora terminoloških kandidatov smo morali razrešiti nekaj temeljnih dilem sodobne leksikologije in terminografije. Osnovna dilema, ali temeljiti na elektronski ali neelektronski obliki jezikovnih podatkov, je bila sicer že razrešena takoj, ko smo se pri pripravi turističnega terminološkega slovarja odločili za korpusni pristop. Kljub temu pa so ostala odprta vprašanja o smiselnosti vključevanja nedokumentirane leksike, in sicer o dopolnjevanju geslovnika s prevajanjem in prirejanjem delov tujih slovarjev in o tvorbi novih terminov zgolj zaradi normativnosti. Skladno s tujimi in domačimi sodobnimi usmeritvami (glej Logar, 2009, 321–326) pri odločitvah dajemo prednost izpričani jezikovni rabi v korpusu turističnih besedil. Dopolnjevanje geslovnika s pomočjo tujih slovarjev in enciklopedij turizma se pojavlja le v primerih, ko turistična stroka presodi, da je tiste izraze, ki se kot tujke pojavljajo v turističnem diskurzu v slovenščini, zaradi njihove uporabnosti smiselno vključiti v slovar. Hkrati se poskušajo oblikovati novi izrazi za določene žargonske izraze,<sup>2</sup> ki so tudi večinoma prevzeti iz angleškega jezika.

Iz korpusa turističnih besedil, ki je bil oblikoslovno označen in lematiziran z orodjem TOTALE (Erjavac, 2006), je bila avtomatsko generirana lista besed (monogramov) in besednih zvez (bigramov in trigramov), urejena po pogostosti od najvišje k najmanjši. Oznake ter leme smo uporabili za izbiro monogramov na osnovi leme, saj predstavljajo slovarske iztočnice. Za iskanje bigramov in trigramov kot terminoloških kandidatov leme niso uporabne, uporabili pa smo oznake, saj smo tako lahko izločili nemogoče kombinacije bigramov ter trigramov. Ker so avtomatsko generirani sezname besed predvsem iz večjih korpusov lahko zelo obsežni in je v kasnejši fazi identifikacije terminoloških kandidatov potrebno tudi ročno luščenje kandidatov, je bilo potrebno te izjemno dolge seznane primerno avtomatsko skrajšati oz. očistiti.

Seznane smo najprej skrajšali tako, da smo jih prečesali s tujimi (angleškim, italijanskim, nemškim) leksikoni. Računalniški algoritem je tako s seznamoma izločil vse besede, ki so se pojavljale v naštetih tujih leksikonih. S to odločitvijo smo se zavestno odpovedali iskanju določenih tujk, predvsem v angleškem jeziku, ki so v slovenskem turističnem diskurzu precej pogoste, nekatere pa celo nimajo prevodne ustreznice v slovenščini (npr. *last minute*, *all-inclusive*). Da bi ne izgubili tega, za

2 O navzočnosti žargonizmov tujega porekla kot o sestavnem delu besedišča različnih strokovnih področjih gl. tudi Makarova, 2009, 115: "Vobšče pri uglubenii v ljubuju oblast' dejatel'nosti, istoričeski svjazannuju s kakim-libo remeslom (osobenno gorodskim) [...] v slovenskom razgovornom jazyke s vysokoj stepenju reguljarnosti vstrečaejsja plast zaimstvovanij, kotorye, preterpev fonetičeskiju i morfoložičeskiju adaptaciju, učastvujuť v formirovanii remeslennogo žargona". ("Na splošno pri obravnavi besedišča neke stroke, ki zgodovinsko izhaja iz obrtniške dejavnosti (še posebej v mestih) [...] v slovenskih kolokvialnih različicah dokaj pogosto naletimo na tujke in sposojenke, ki po končani glasoslovni in oblikoslovni adaptaciji bistveno prispevajo k slovenskemu obrtniškem žargonu").



**Sl. 1: Turistični vodniki po Slovenski Istri (foto: M. Koderman).**  
**Fig. 1: Tourist guidebooks to Slovene Istria (photo: M. Koderman).**

slovar gotovo pomembnega leksikalnega gradiva, smo seznam kasneje prečesali še s slovenskim, italijanskim in nemškim leksikonom, ter tako pridobili seznam angleških besed, ki je bil uporaben z dveh vidikov:

- v njem smo lahko z ročnim pregledovanjem identificirali tiste izraze, ki se kot tujke pojavljajo v turističnem diskurzu v slovenščini;

- seznam besed je uporaben tudi kot baza prevodnih ustreznice za terminološki slovar, ki bo po svoji zasnovi dvojezičen, slovensko-angleški; omenjene izraze in njihovo ustreznost lahko preverimo v korpusu in drugih dostopnih referenčnih virih.

Seznam smo dodatno omejili tudi s tem, da smo določili mejo števila pojavitev, do katerega bomo gradivo, pridobljeno iz korpusa, še pregledovali. Logar Berginc in Vintar (2008) na primeru KoRP-a (Korpusa odnosov z javnostmi) sicer ugotavljata, da načelna odločitev, ki velja za splošne slovarje in mejo pojavitve v korpusu postavlja relativno visoko (npr. na 100 pojavitev v 100 ali večmilijonskem korpusu), za terminološke slovarje ne velja, saj je za strokovni jezik značilno, da je razmerje med specifičnostjo terminološkega poimeno-

vanja in pogostnostjo v korpusu pogosto prej obratno kot prenosorazmerna. Liste besed bi bilo tako v našem primeru treba pregledati v celoti: ker pa bi lista besed iz več kot 30 milijonskega korpusa besed potrebovala izjemno velik obseg vključevanja dodatnih sodelavcev pri projektu (seznam monogramov, pridobljenih iz korpusa, je obsegal 500.000 različnic), smo se tu vendarle morali odločiti pragmatično: meja pogostosti je bila tako postavljena na nizko mejo treh pojavitev (ročno pregledane so bile vse različnice, ki se v korpusu pojavljajo vsaj trikrat).

Nato smo za vsak seznam pripravili še nabore besed, ki naj jih program avtomatsko izloči, saj smo zanje lahko z gotovostjo predpostavljali, da ne predstavljajo kandidatov za termine. Take besede so bile določeni vezniki (*in, ter, ker, če, da, ampak ...*), osebni, kazalni, celostni in oziralni zaimki; predlogi (*v, k, h, od, o, z, s, za*), členki (*tudi, že, še, le, no, mar ...*), naklonski izrazi (*naj, lahko ...*); pomožni glagoli ter prislovi (*zdaj, tedaj, danes, jutri, včeraj, odslej, odtlej, poslej, tukaj, tam, prvič, drugič, tretjič*). Očiščene sezname smo nato sodelavci pri projektu ročno pregledali in po svoji presoji v njih

označili kandidate za termine. Pri tem smo upoštevali sprejeto odločitev projektne skupine, da bo slovar vseboval predvsem terminologijo turistične industrije in infrastrukture (ali "turizem v ožjem pomenu besede"), poleg tega pa tudi nekatere termine, ki izvirno prihajajo z drugih področij, npr. zgodovine, arhitekture ali geografije, ki pa so s turizmom kot sestavljeno dejavnostjo neločljivo povezani. Ker je glede na naravo turizma kot dejavnosti teh področij kar nekaj, bi lahko postal slovar hitro preobsežen, zato smo se odločili s teh področij zajeti res samo tiste termine, ki so za turistični diskurz tako značilni in pogosti, da je upravičena njihova vključitev v slovar; nismo pa v slovar vključevali vseh tistih, ki bi glede na področje smiselno spadali v okvir terminoloških slovarjev z omenjenih področij, npr. v arhitekturni terminološki slovar. Dodati je treba, da smo v korpusu iskali ne le termine, temveč tudi terminološke kolokacije, saj predstavljajo pomemben del strokovnega izrazja.

Odločitev, katere leksikalne enote vključiti v slovar in katere ne, je seveda "mehka" oziroma v določeni meri subjektivna. Kot opozarjata Logar Berginc in Vintar (2008, 14), so "korpusi strokovnih besedil razrahljali več mej: mejo med terminološko in neterminološko leksiko, mejo, do katere še govorimo o večbesedni poimenovalni enoti in čez katero je že prostor kolokacij, ter mejo, ki določa, ali gre za termin področja, ki ga obravnavamo, ali ne".

Zato smo v primeru dvoma, ali naj določeno leksikalno enoto označimo kot termin ali ne, izvedli še korpusno poizvedbo: preverili smo pogostost besede, njeno besedilno okolje, besedna skica, ki jo nudi programska oprema SketchEngine, nam je pokazala leksemske povezave, za mnenje smo povprašali tudi zunanje strokovnjake s področja turizma, ki sodelujejo pri projektu (Slovenska turistična organizacija kot partner in sofinancer projekta ter Fakulteta za turistične študije Turistica Univerze na Primorskem).

### OBLIKOVANJE GESELSKEGA ČLANKA

Pri oblikovanju geselskega članka je bilo treba sprejeti odločitev o tem, katere vrste informacij bomo v članek vključili in na kakšen način. Poleg standardnih slovarskih informacij, kot so npr. oznake za besedno vrsto, morebitni fonetski zapis, definicija in prevodna ustreznica v tujem jeziku (pri večjezičnih slovarjih), smo se tu odločali še za določene oznake, ki so specifične za terminološki slovar. Geselski članek smo strukturirali v obvezna in dodatna ali opcijska polja. Med obvezna polja smo tako poleg seveda iztočnice vključili še oznako besedne vrste (brez navajanja slovničnih posebnosti), definicijo oziroma definicije, številčene po vrstnem redu, ter prevodno ustreznico v angleškem jeziku.

Med dodatna opcijska polja smo vključili:

- kvalifikator za turistično zvrst, povzeto in prilagojeno po Tezavru turizma in priložnostnih dejavnosti

Svetovne turistične organizacije (celoten seznam zvrsti je objavljen v Mikolič et al., 2008);

- kvalifikator za področje družbene dejavnosti, in sicer za področja družbene dejavnosti, ki ne sodijo v okvir turistične industrije kot turizma v ožjem pomenu besede, pač pa v okvir turizma kot sestavljene dejavnosti, kot npr. zgodovina umetnosti ali geografija, ki pa vsebujejo lekseme, ki so v tolikšni meri značilni za turizem, da jih lahko štejemo za turistično terminologijo;

- pogoste terminološke kolokacije, torej besedne zveze, ki jih ne moremo uvrstiti pod iztočnice, so pa reprezentativne za turistični diskurz;

- sopomenke, kjer obstajajo;

- kazalke na druga, pomensko sorodna gesla, z dvema oznakama: oznako PRIM. ("primerjaj") za lekseme, med katerimi obstaja skoraj sinonimija ali navidezna sinonimija, ter GL. ("glej") za druga pomensko sorodna gesla;

- normativno oznako, s čimer smo želeli odgovoriti na občuteno potrebo turistične stroke in industrije po pomensko jasni in normativno ustrezni terminologiji.

Strukturo geselskega članka navajamo v nadaljevanju:

Geslo / slovnična oznaka / *normativna oznaka* / kvalifikator (tur. zvrst) / kvalifikator (področje družbene dejavnosti) / definicije (številčene po vrstnem redu, zaporedne številke v krepkem tisku) / pogoste kolokacije / sopomenke / kazalke na druga gesla (GL. in PRIM.)

Ustreznica v angleškem jeziku (ločena vrstica)

### SKLEP

Korpusni pristop k luščenju turistične terminologije se pokaže kot dovolj objektivna in uporabna metoda za identifikacijo terminoloških kandidatov, pri čemer je pomembna jasno utemeljena klasifikacija besedil, zajetih v področni korpus. Dopolnjevanje geslovnika s pomočjo tujih slovarjev in enciklopedij turizma se pojavlja le izjemoma, zaradi manjkajočega termina v slovenskem jeziku za tiste izraze, ki se kot tujke pojavljajo v turističnem diskurzu v slovenščini, in za žargonske izraze.

Slovar bo na voljo v elektronski in tiskani obliki. Geselski članki se urejajo s sodobno in po zasnovi zelo prožno in uporabniku prijazno programsko opremo, ki nam jo je v uporabo ponudilo podjetje Amebis. V elektronski verziji slovarja bo mogoče iskati slovenske in angleške iztočnice, pri čemer bodo slovenske iztočnice obdelane v geselskem članku, angleškim iztočnicam pa bo dodana samo prevodna ustreznica v slovenskem jeziku. Tiskana različica slovarja pa bo v prvem delu obsegala slovenske iztočnice, obdelane v geselskem članku, drugi del slovarja pa bo vseboval seznam angleških iztočnic s prevodnimi ustreznici v slovenskem jeziku. Taka zasnova slovarja je utemeljena v ciljih aplikativnega raziskovalnega projekta Turistični terminološki slovar, ki poteka v sodelovanju s Slovensko turi-

stično organizacijo. Namen aplikativnega projekta je bil namreč prav odzvati se na potrebe turistične stroke in industrije v Sloveniji po nedvoumni terminologiji in kakovostnem strokovnem sporazumevanju predvsem v domačem, pa tudi mednarodnem okolju. Elektronska oblika obeh orodij, korpusa in slovarja, omogoča, da se lahko sproti dopolnjujeta. Sprotno ažuriranje teh orodij, ki so na eni strani nepogrešljiva za razvoj nekega področja, na drugi strani omogoča sledenje hitri rasti tega področja, kar za turizem nedvomno velja.

Sklenemo lahko, da so odločitve, do kolikšne mere posegati na različna družbena (pod)področja, ki se vključujejo v turizem, vedno težavne in tudi bolj ali manj subjektivne, še posebej zato, ker je turizem tako široko in sestavljeno področje. Če bi se v primeru tursitičnega

slovarja odločali zajeti samo termine s področja turizma v ožjem pomenu besede, tj. turistične industrije, bi bil izbor ožji in zato lažji. Ker pa je bil naš namen na turizem pogledati širše in torej zajeti tudi različna (pod)področja, ki se vključujejo v turistično dejavnost, bi lahko postal slovar hitro preobsežen, zato je bilo v procesu identifikacije terminoloških kandidatov potrebno ob podatku o številu pojavitev v korpusu vsak izraz analizirati tudi z vidika njegovega pomena, ki ga ima za specifično strokovno področje (ali več področij) na eni strani in za področje turizma v širšem pomenu besede na drugi strani. Tudi zaradi tega je kvalifikator, ki sporoča (pod)področje rabe nekega termina, pomemben element geselskega članka.

## IDENTIFYING TERMS FOR A DICTIONARY OF TOURIST TERMS

Vesna MIKOLIČ

University of Primorska, Faculty of Humanities Koper, SI-6000 Koper, Titov trg 5  
e-mail: vesna.mikolic@fhs.upr.si

Ana BEGUŠ

University of Primorska, Faculty of Humanities Koper, SI-6000 Koper, Titov trg 5  
e-mail: ana.begus@fhs.upr.si

### SUMMARY

*Dictionaries of terms are characteristic of highly developed expert fields and enable linguistically appropriate expert communication, they suggest Slovene terms for new subjects and terms within quickly developing or new expert fields, reflect the spirit and culture of a nation (if they are based upon local material!) as well as political tendencies (Humar, 2004, 23). At the same time the multilingual terminological dictionaries enable international communication within new social and political circumstances. At this point, surely, the way a terminological dictionary is conceived and formed, should be put into consideration. Logar (2009, 321) refers to three phases in the creation of a dictionary of terms, namely: determining the texts suitable for term acquisition; defining the standards for including words into the dictionary; defining information included within a dictionary article. The aim of the present paper is to show the theoretical definitions and practical solutions important for all three phases of the developing dictionary of tourist terms with an emphasis on term identification. Editing a dictionary of tourist terms was based on corpus approach. Firstly, the paper focuses on selection and classification of texts for the TURK corpus of tourist texts, which currently includes approximately 30 million words. The corpus was used as a basis for an automatically generated list of words (monograms) and phrases (bigrams and trigrams) arranged from most to least frequent. The further procedure of selecting tourist terms was performed on the basis of preliminary sorting of tourist texts in the Corpus according to the type of tourism and the included social sphere. It was precisely the social sphere, which is expressed in the dictionary article as a qualifier of an individual term, that has proven to be the most relevant when making a selection of tourist terms.*

**Key words:** tourism, terminology, dictionary of terms, candidate term, specialised corpus, field-specific texts, functional typology, dictionary article

## LITERATURA

- Erjavec, T. (2006):** Multilingual tokenisation, tagging, and lemmatisation with totale. V: Abstracts / 9th INTEX/NOOJ Conference, Belgrade, Serbia, June 1–3, 2006. Belgrade, Faculty of Mathematics, University of Belgrade.
- Goeldner, C. R., Brent Richie, J. R. (2003):** Tourism. Principles, Practices, Philosophies. Hoboken, New Jersey, John Wiley & Sons, Inc.
- Gorjanc, V., Logar, N. (2007):** Od splošnih do specializiranih korpusov – načela gradnje glede na njihov namen. V: Orel, I. (ur.): Razvoj slovenskega strokovnega jezika. Obdobja, Metode in zvrsti, 24. Ljubljana, Filozofska fakulteta, Oddelek za slovenistiko, Center za slovenščino kot drugi/tuji jezik, 637–650.
- Gorjanc, V. (2010):** Terminološko načrtovanje in upravljanje terminologije. Slavistična revija, 58, 1. Ljubljana, 95–104.
- Humar, M. (ur.) (2004):** Terminologija v času globalizacije: zbornik prispevkov s simpozija Terminologija v času globalizacije. Ljubljana, Znanstvenoraziskovalni center SAZU, Založba ZRC.
- Humar, M. (2009):** Položaj in prihodnost slovenske terminologije in terminografije. V: Ledinek, N., Žagar Karer, M., Humar, M. (ur.): Terminologija in sodobna terminografija. Ljubljana, Založba ZRC - ZRC SAZU, 41–46.
- Ledinek, N., Žagar Karer, M., Humar, M. (ur.) (2009):** Terminologija in sodobna terminografija. Ljubljana, Založba ZRC - ZRC SAZU.
- Logar, N. (2009):** Korpusi v terminografiji: umik potrebe po introspektivni presoji. V: Ledinek, N., Žagar Karer, M., Humar, M. (ur.): Terminologija in sodobna terminografija. Ljubljana, Založba ZRC SAZU, 319–328.
- Logar Berginc, N., Vintar, Š. (2008):** Korpusni pristop k izdelavi terminoloških slovarjev: od besednih seznamov in konkordanc do samodejnega luščenja izrazja. Jezik in slovstvo, 53, 5. Ljubljana, 3–17.
- Makarova, I. (2009):** Problemy slovenskoj jazykovoj situaciji i osobennosti ljubljanskoj razgovornoj reči. Moskva, Filologičeskij fakul'tet MGU im. M.V. Lomonosova.
- Mikolič, V. (2007):** Tipologija turističnih besedil s poudarkom na turističnooglaševalskih besedilih. Jezik in slovstvo 52, 3–4. Ljubljana, 107–116.
- Mikolič, V. (2009):** Specializirani jezikovni korpusi in funkcijska zvrstnost. V: Stabej, M. (ur.): Infrastruktura slovenščine in slovenistike. Obdobja, Simpozij. Ljubljana, Znanstvena založba Filozofske fakultete, 257–263.
- Mikolič, V., Beguš, A., Dukič, D., Koderman, M. (2008):** Vpliv namembnosti korpusa na označevanje besedilnega gradiva za "Večjezični korpus turističnih besedil". V: Erjavec, T., Žganec Gros, J. (ur.): Zbornik Šeste konference Jezikovne tehnologije, 16. do 17. oktober 2008, zbornik 11. mednarodne multikonference Informacijska družba – IS 2008, zvezek C (Informacijska družba). Ljubljana, Institut Jožef Stefan, 60–64.
- Stramljič Breznik, I. (2003):** Besedotvorna tipologija novonastalega besedja s področja mobilne telefonije. Ljubljana, Slavistično društvo Slovenije, 105–118.
- Thesaurus on Tourism and Leisure Activities (2001).** World Tourism Organization (WTO), Secretariat of State for Tourism of France.
- Verdonik, D., Žgank, A., Pisanski Peterlin, A. (2008):** Validacija označevanja diskurzivnih označevalcev v korpusih Turdis-2 in BNSlint. V: Erjavec, T., Žganec Gros, J. (ur.): Zbornik Šeste konference Jezikovne tehnologije, 16. do 17. oktober 2008, zbornik 11. mednarodne multikonference Informacijska družba – IS 2008, zvezek C (Informacijska družba). Ljubljana, Institut Jožef Stefan, 29–32.
- Vintar, Š. (2008):** Terminologija: terminološka veda in računalniško podprta terminografija. Ljubljana, Znanstvena založba Filozofske fakultete, Oddelek za prevajalstvo.