

Maja Bratanić
Ana Ostroški Anić
Institut za hrvatski jezik i jezikoslovlje
Ulica Republike Austrije 16, HR-10000 Zagreb
mbratan@ihjj.hr, aostrosk@ihjj.hr

PEDAGOŠKI PRISTUP KORPUSNO UTEMELJENOJ IZRADBI KOLOKACIJSKOGA RJEČNIKA STRUKOVNOGA NAZIVLJA

U radu se opisuje postupak izradbe specijaliziranoga i u tipološkom smislu inovativnoga terminološkoga kolokacijskog rječnika. Na temelju namjenski odabrana uzorka engleskoga zrakoplovnog jezika – s ciljem da se precizno zahvati nazivlje i vokabular jezika struke za pilote i kontrolore letenja na globalno propisanoj razini – izrađen je opsežan popis (od gotovo 10 000) engleskih zrakoplovnih naziva i njihovih tipičnih leksičkih sveza. U njegovoj su izradbi sudjelovali studenti viših godišta studija aeronautike uz razmjerno jednostavnu korpusnolingvističku podršku. Ovakva djelomice računalno podržana metoda izlučivanja reprezentativnih terminologiziranih kolokacija na konkretnom se materijalu pokazala djelotovornijom od automatiziranih pristupa, koji ne uzimaju u obzir sve kriterije odabira terminološki relevantnih leksičkih sveza, a i u odnosu na tradicionalne leksikografske i terminografske metode. Dobivena leksička građa izravno je primjenjiva kao “lijeva strana” dvojezičnih rječnika zrakoplovnoga nazivlja, zanimljiva za daljnja leksikološka istraživanja, ali i potencijalno iskoristiva u aktivnom usvajanju (stranoga) strukovnoga nazivlja.

Uvod

Izradba rječnika zrakoplovnih naziva i kolokacija *Aviation English Terms and Collocations (An Alphabetical Checklist)* (Bratanić et al. 2010) bila je motivirana obvezom koju je uvela Međunarodna organizacija za civilno zrakoplovstvo (ICAO), a kojom se propisuje da svi djelatni piloti i kon-

trolori zračnoga prometa u svijetu najkasnije do ožujka 2011. iz sigurnosnih razloga moraju dokazati odgovarajuće poznavanje engleskoga jezika kao dogovorenoga međunarodnog jezika zrakoplovstva. Tim se zahtjevima bavi opsežna stručna literatura pa se propisano pretpostavljeno jezično znanje opisuje razmjerno detaljno razrađenim kriterijima za različite oblike jezične kompetencije. Lako je uočiti, međutim, da se leksičko znanje — ono što se u strukovnom jeziku uglavnom može opisati kao poznavanje vokabulara — uglavnom implicitno i načelno samorazumijeva spominjanjem brojnih, ali široko zacrtanih leksičkih domena. Rad na leksičko-leksikografskom projektu koji ovdje opisujemo suočio se, stoga, sa zadatkom da se ta nedovoljno definirana leksička komponenta egzaktno utvrdi i sustavno popiše.

Izradba specijaliziranoga rječnika zrakoplovnih naziva i kolokacija polazi od pretpostavke da će se propisani leksički fond zrakoplovnoga engleskog jezika najdjelotvornije utvrditi na ciljanom korpusu i egzaktnim kriterijima.

Svrha je rada na ovom projektu bila ipak višestruka. U prvom se redu htjelo na osnovi korpusa izraditi iscrpan popis relevantnih leksičkih kombinacija ili kolokacija koji će biti upotrebljiv u pedagoške svrhe s gledišta stranoga jezika struke, a zatim ustanoviti leksikografski potencijal tako dobivene građe za rječnik zrakoplovnih kolokacija, odnosno za “lijevu stranu” zamišljenoga dvojezičnog specijaliziranog rječnika. Na tom su projektu angažirani apsolvirani ili netom završeni studenti aeronautike na Zavodu za aeronautiku Fakulteta prometnih znanosti jer je projekt počeo kao svojevrsna leksikografsko-terminografska vježba s ciljem da se obuhvati vokabular kojim studenti aeronautike moraju (većim dijelom aktivno) raspolagati do stjecanja svih profesionalnih dozvola.

Zanimljivo je, nadalje, iz perspektive nastave stranoga jezika struke, bilo i utvrditi je li takav metodološki zahtjevan leksičko-leksikografski poduhvat moguće koncipirati kao pedagoški zadatak te koliko su, naposljetku, rezultati takva rada i praktično i teorijski primjenjivi u nastavi strukovnoga jezika. Ovo posljednje pitanje, međutim, nadilazi okvire ovoga rada i predmetom je drukčije vrste analize, no budući da je bilo i jednim od poticaja za ovako koncipiran rad na rječniku, nužno je i to spomenuti.

Za korpus je izabrana građa propisana kao obvezna za stjecanje međunarodno priznate dozvole prometnoga pilota (ATPL) i kontrolora zračnoga prometa, tj. za sve sudionike u međunarodnom zračnom prometu. U korpus je stoga uvrštena propisana ispitna literatura, banka pitanja, temeljni dokumenti za kontrolu zračnog prometa i manji broj priručnika nezaobilaznih za stjecanje odgovarajućih dozvola.

Kolokacije

Potrebno je ukratko objasniti radnu definiciju i opseg pojma leksičke kolokacije na koju se rječnik metodološki oslanjao.

Premda se kolokacija kao svojevrsna idiosinkratična leksičko-semantička sveza proučava više od pedeset godina, istraživanja su se bitno množila s napretkom korpusne lingvistike i sve većom dostupnošću odgovarajućih metoda zahvaljujući pristupačnim programima za korpusnu analizu. Već je J. Firth, koji je tu leksičku pojavu i imenovao,¹ uočio da je kolokacijsko supojavljivanje dvaju leksičkih elemenata statistički mjerljivo.

Tradicionalno se o kolokacijama razmišlja kao o leksičkim kombinacijama dvaju leksema u sintagmatskom supojavljivanju statistički češćem no što je to slučaj s bilo kojim drugim leksemima koji se mogu naći u najčešćim kolokacijskim supostavama (imenica + imenica, pridjev + imenica, glagol + imenica, glagol + prilog i pridjev + prilog). Za razliku od značenja fraza (engl. idioma), značenje takvih dvorječnih sveza za izvorne je govornike transparentno te će oni najčešće bez problema prepoznati “ispravnost” takvih sklopova. Sposobnost razlikovanja “točnih” od “netočnih” leksičkih sveza po našem se mišljenju više temelji na iskustvu izvornih govornika negoli na njihovoj intuiciji jer su kolokacijske veze često arbitrarne, a introspekcija se i u istraživanjima kolokacija pokazala nepouzdanim kriterijem zato što je znanje jezika i u izvornih govornika ograničeno, a predrasude lako moguće (v. npr. Greenbaum 1974:83, Krishnamurthy 2000:32–33). Intuicija pak neizvornih govornika pokazuje se i na visokoj razini poznavanja stranoga (u ovom slučaju kolokacijski vrlo tvorbenoga engleskoga) nedostatnom u procjeni valjanosti kolokacija.

I tu smo tezu željeli testirati u sastavljanju građe našega rječnika, polazeći prije svega od pretpostavke da samo reprezentativno izabran korpus može razotkriti vjerojatnost semantičkih obrazaca koja nadilazi intuiciju i govornu praksu pojedinačnih izvornih govornika (Stubbs, 2001:153).

Nema jedinstvene definicije kolokacija u jezikoslovnoj literaturi, a u ovom je radu ona pomaknuta i preko granica uobičajenijih određivanja. U kontekstu jezika struke s jedne strane, odnosno korpusne i računalne leksikografije i terminografije s druge, produktivno je i pragmatično, po našem mišljenju, raspon kolokacija proširiti na svaku arbitrarnu kombinaciju riječi koja se u reprezentativnome korpusu javlja češće od očekivanoga. Sličan fleksibilan stav dijele i John Sinclair (2004), Dirk Siepmann (2005) te neki autori koji govore o terminološkim kolokacijama kao o svojevrsnoj podvrsti leksičkih sveza.

¹ *Papers in Linguistics* 1934–1951, Oxford : Oxford University Press, 1957.

Posljednjih godina razni autori sve češće upozoravaju na važnost kolokacija ne samo u širem leksikografskom kontekstu, već i u izradbi strukovnih odnosno terminoloških rječnika (među prvima, npr. Bergeholtz i Tarp 1995). Uobičajeno se fenomen kolokacija povezivao više s općim jezikom nego sa strukovnim, no problemi koje kolokacije sa sobom nose, a koji se najlakše manifestiraju u prevođenju, podjednako se primjećuju i u strukovnim jezicima. No, u praksi se kolokacijske sveze specifične za strukovne jezike rijetko sustavno leksikografski prezentiraju, a u terminološkim se zbirkama i bazama također obično ne bilježe ni iscrpno ni sustavno. U vrijeme rada na ovom projektu nije nam bio poznat ni jedan kolokacijski strukovni rječnik pa je ovaj priručnik izvorni prilog terminografskoj metodologiji. Ni konvencionalni terminološki rječnici ne nude razrađene modele jer su u njima pored oskudne zastupljenosti kolokacija uočljive i nedoumice vezane uz njihovu prezentaciju.

U istraživanju koje je prethodilo izradi priručnika *Aviation English Terms and Collocations* nije se povlačila oštra granica između višerječnih kolokacija i višerječnih termina, odnosno između leksičkih sveza sa snažnim terminološkim potencijalom i etabliranih višerječnih naziva. U ovoj je pristupu taj pragmatični okvir još radikalnije proširen i na kratice, odnosno njihove razriješnice, zbog njihove izuzetne učestalosti u stručnom zrakoplovnom jeziku i visokoga stupnja njihove terminologiziranosti.

Može se, dakako, postaviti pitanje o razgraničenju između kolokacija u tradicionalnom smislu i višerječnih izraza. Marie-Claude L'Homme (2009b) dvojí i o tome može li se uopće pojam kolokacija iz općega jezika prenijeti u strukovni jezik i proširiti ga na fenomen koji ona naziva »specijaliziranim leksičkim kombinacijama«. Autorica smatra da — premda je u jeziku struke praktično, pa i potrebno, uzeti u obzir višerječne leksičke sveze labavije od tradicionalno shvaćenih kolokacija — valja praviti razliku između tih dvaju pojmova kao fenomena jezika struke odnosno općega jezika. Ona nadalje drži da se ni prototipske kolokacije ne ponašaju isto u općem jeziku i u strukovnom jeziku te takvu tezu brani procjenom da se leksemi koji se učestalo javljaju u »specijaliziranim leksičkim kombinacijama« mogu kombinirati s cijelim skupinama semantički srodnih riječi, što umanjuje nepredvidljivost kolokiranja specifičnu za tradicionalno shvaćene kolokacije.

Premda se ove postavke do stanovite mjere mogu braniti s teorijskih pozicija, ne vidimo prava razloga da se, uzimajući u obzir korisničke potrebe, leksičke sveze učestale u jezicima struka u rječnicima ne prezentiraju na isti način kao i kolokacije. Hoće li se one nazivati kolokacijama ili višerječnim specijaliziranim izrazima može biti predmetom rasprave, no

polazeći od iskustva neizvornoga govornika ta se razlika čini nedovoljno relevantnom kao leksikografsko-terminografska činjenica.

Podatak da se kolokati mogu supojavljivati s različitim semantički srodnim leksemima ne povećava bitno njihovu predvidljivost u konkretnom jeziku struke.

Cilj projekta, korpus i metoda

Cilj je projekta bio analizom korpusa sastaviti što iscrpniji popis zrakoplovnih naziva i leksičkih sveza koje bi se mogle smatrati punopravnim nazivima ili terminologiziranim izrazima te tako ujedno i kandidatima za budući englesko-hrvatski dvojezični zrakoplovni rječnik. Tako bi dobivena građa mogla poslužiti kao “lijeva strana” bilo kojega dvojezičnog rječnika zrakoplovnoga nazivlja s engleskim kao polaznim jezikom. Lako je utvrditi da brojni izrazi koji se susreću u literaturi nisu na odgovarajući način ili nikako zabilježeni u postojećim (jednojezičnim i višejezičnim) zrakoplovnim rječnicima.

Za analizu korpusa i odabir relevantnih kolokacija sastavljači rječnika služili su se programom *Wordsmith*, verzijom 4.0. iz 2007. godine. S obzirom na nepretencioznu početnu namjenu projekta, korpus nije bio označen. Izrađeni su čestotni i abecedni popisi a potom su iz konkordancija, prikazanima tako da istaknu lijevi ili desni kontekst stožernice, izabirane kolokacije.

U radu se potvrdila pretpostavka da je bez obzira na često isticanu nepredvidljivost i neprozirnost kolokacija i neizvornu govorniku s prosječnim poznavanjem struke u velikoj mjeri moguće intuitivno prepoznati — ne toliko valjanost kolokacija i višerječnih stručnih izraza koliko njihove granice. Na građi prikupljenoj u toj fazi — koju su savjesnim pretraživanjem korpusa prikupili apsolvenci ili netom završeni inženjeri aeronautike s dobrim operativnim znanjem engleskoga jezika i temeljnim poznavanjem načela korpusne leksikografije — trebalo je provesti “čišćenje” i odabir relevantne građe.

Tako su dobivenu građu sredila tri urednika rječnika, uglavnom reduciranjem tzv. nepravilnih kolokacija, opetovanom provjerom u korpusu i intervencijama u vezi s odabirom osnovne riječi i njezinom gramatičkom kategorizacijom. Pri odabiru “pravih” kolokacija najbitnija je bila, opet u nekoj mjeri arbitrarna, procjena o stupnju njihove terminologiziranosti.

Konačan popis sadrži oko 10.000 kolokacija. Već se i grubom procjenom može ustanoviti da znatno pretežu dvorječne kolokacije, no i da je

zbog specifične prirode ovoga leksika i višerječnih kolokacija više nego u općem jeziku. Gramatičke kolokacije nisu uključene u popis, osim iznimno kao dijelovi većih višerječnih leksičkih sveza.

Prikaz građe

Kolokacije odnosno leksičke sveze organizirane su oko osnovne riječi. Budući da u leksičkim kolokacijama, posebno onima u jeziku struke, nije uvijek ni u konceptualnom ni u semantičkom smislu lako odrediti koja je riječ osnovna, sve se kolokacije navode pod svim sastavnim elementima. Važniji je razlog za tu odluku bio i to što se smatralo da će ovako priređena građa korisniku biti znatno dostupnija. Kako bi se osiguralo da se sve kolokacije navedu na odgovarajućim mjestima u popisu, izrađen je i poseban računalni program.

53	
→ colour vision	
coloured <i>adj</i>	
coloured airways	
coloured flag	
coloured flag system	
column <i>n</i>	
column of air	
control column	
combined <i>adj</i>	
combustibility <i>n</i>	
combustion <i>n</i>	
annular combustion chamber	
can-type combustion chamber	
combustion air	
combustion air fuel ratio	
combustion air intake	
combustion blower	
combustion chamber	
combustion chamber inlet	
combustion chamber intake	
combustion chamber pressure	
combustion chamber temperature	
combustion heater	
combustion process	
combustion section	
combustion system	
combustion temperature	
complete combustion	
incomplete combustion	
internal combustion	
internal combustion engine	
combustor <i>n</i>	
	commencement <i>n</i>
	commencement point
	commercial <i>adj</i>
	commercial aeroplane
	commercial aircraft
	commercial airline
	commercial air transport
	commercial air transport operation
	commercial aviation
	commercial flight
	commercial helicopter
	commercial operation
	commercial pilot licence [CPL ¹]
	commercial transport aircraft
	commercial transport category
	international commercial aviation
	non-scheduled commercial flight
	scheduled commercial service
	commission <i>n</i>
	Air Navigation Commission [ANC]
	Radio Technical Commission for Aeronautics [RTCA]
	committee <i>n</i>
	Air Coordinating Committee [ACC ¹]
	common <i>adj</i>
	common digitizer
	common frequency
	common language
	common point
	communication <i>n (often pl)</i>
	AFTN communication centre
	aircraft communications equipment

1. Prikaz dijela stranice iz *Aviation English Terms and Collocations*

Glavne se natuknice (jednorječne i po potrebi višerječne) nižu abecednim redom. Slijedi gramatička odrednica, a područna se donosi kada pripadnost koje sveze određenom zrakoplovnom području po procjeni sastavljača nije sama po sebi razumljiva.

Višerječne se sveze i same donose kao osnovne natuknice onda kad za to postoje konceptualni razlozi, kao u sljedećem slučaju:

holding pattern

holding pattern airspace
holding pattern airspeed
holding pattern entry
holding pattern size
holding pattern template
designated holding pattern
non-standard holding pattern
standard holding pattern

ili, kao u primjeru *horse latitude*, *squirrel cage* i slično, gdje bi razdvajanje na sastavne dijelove zahtijevalo uvođenje osnovne natuknice *horse* ili *squirrel* koje se izolirane ne mogu smatrati zrakoplovnim nazivima ili leksemima općega jezika relevantnima za zrakoplovni kontekst. Takve se kolokacije mogu u rječniku naći i kao samostalne kolokacije, ali i pod elementom koji pripada zrakoplovnom nazivlju u užem ili širem smislu (*latitude*, *cage*).

Slično se postupa i s kolokacijama koje sadrže vlastito ime poput *Faraday's Law*, *Gay Lussac's Law*, *Kepler's Laws* itd.

Homografi su razdvojeni u zasebne natuknice kad je riječ o dvama različitim pojmovima:

lead¹ n

low lead [LL]
low lead fuel
spongy lead
tetraethyl lead [TEL]

lead² n

inertial lead vertical speed indicator [ILVSI]

PCU¹ abbr power control unit

PCU² abbr propeller control unit

Najčešće su se nedoumice javljale u vezi s tim treba li osnovnu riječ, a tu je mahom riječ o oblicima na *-ing*, dakle prezentskim participijima, kategorizirati kao pridjeve ili kao imenice. Budući da je to čest leksikografski problem, koji je ponekad uputno razriješiti kompromisom, slijedili smo praksu nedavno temeljito prerađenoga i autoritativnoga *Shorter Oxford English*

Dictionary, na koji smo se oslanjali i u rješavanju ortografskih pitanja vezanih poglavito uz složenice, ali i druge pravopisne odluke.

Radi ekonomičnosti prezentacije uz pojmovno i tvorbenu srodne glavne natuknice navode se gramatičke odrednice za sve vrste riječi u kojima se javljaju.

anti-icing *n/adj*

anti-icing equipment
anti-icing facility
anti-icing fluid
anti-icing holdover time
anti-icing measure
anti-icing pad
anti-icing system
bleed air anti-icing
electrical anti-icing
engine anti-icing system
engine intake anti-icing
leading edge anti-icing
thermal anti-icing

ili

bias *n/adj*

bias spring
bias voltage
clock bias
decision bias
sea bias *nav*

Kad je procijenjeno da bi takav postupak mogao zbuniti korisnika, zadržana je podjela prema odgovarajućim gramatičkim kategorijama.

aerial *adj*

aerial advertisement
aerial loading unit [ALU]
aerial rotation
aerial screening
aerial work

aerial *n*

dipole aerial
gyro-stabilized aerial array
horizontal aerial
ILS glide path aerial
localizer aerial
loop aerial
Marconi aerial

radar aerial
radio aerial
receiver aerial
sense aerial

Budući da je osnovni kriterij prezentacije kolokacija bila prozirnost i lako snalaženje u bogatoj građi, poteškoće u vezi s makrostrukturom i mikrostrukturom rječnika uglavnom su bile vezane uz ne uvijek jednostavnu odluku o tome kojoj gramatičkoj vrsti riječi pripada neki od elementata kolokacije. Npr., treba li kolokaciju *stack pipe* navesti pod imenicu *stack* ili glagol *stack*, je li u slučaju *scratch pad scratch* glagol, imenica ili čak pridjev.

Trebalo je odlučiti i, kako je spomenuto, treba li koja kolokacija biti klasificirana kao samostalna natuknica ili abecedirana pod pojedinačnim sastavnim elementima (poput *light chop*, *bear paw*) kad osnovna riječ nema odgovarajuću terminološku vrijednost.

U nekim je slučajevima trebalo provesti pravu semantičku analizu i procijeniti je li, primjerice, u kolokaciji *squat switch*, primarno značenje leksema ono imenice *squat* 'čučanj' jer je riječ o prekidaču-senzoru koji reagira kad zrakoplov leži ('čuči') na zemlji, a ne o 'niskom i zdepastom' prekidaču.²

Usporedba s drugim rječnicima

Aviation English Terms and Collocations u tipološkom je smislu inovativna vrsta specijaliziranoga kolokacijskog terminološkog rječnika u potpunosti utemeljena na korpusu.

Usporedbom dobivene građe s nazivima i kolokacijama koji se nalaze u dvama opsežnijima jednojezičnim engleskim rječnicima zrakoplovstva (*The Cambridge Aerospace Dictionary* i *Dictionary of Aviation*) te s jednim englesko-hrvatskim rječnikom (*Englesko-hrvatski zrakoplovni rječnik*), zaključeno je da se ovom metodom dobilo znatno više kolokacija nego tradicionalnom metodom sastavljanja rječnika, koja se često temelji na kompilaciji već raspoloživih leksikografskih priručnika. To se može pokazati na primjeru tri važna zrakoplovna pojma — *airline*, *airspace* i *runway*.

² U takvim i mnogim drugim nedoumicama od neprocjenjive nam je koristi bila ekspertiza profesora Waylesa Brownea sa sveučilišta Cornell, SAD.

AE Terms and Collocations (Bratanić et al. 2009)	The Cambridge Aerospace Dictionary (Gunston 2004)	Dictionary of Aviation (Crocker 2007)	Englesko-hrvatski zrakoplovni rječnik (Eleršek 2008)
airline <i>n</i> airline crew airline deregulation act airline operations centre [AOC ¹] airline operator airline passenger airline pilot airline radio operator airline transport pilot licence [ATPL] commercial airline hub-and-spoke airline system	airline	airline airline representative airline security area Airline Transport Pilot's Licence	airline airline cargo airline charter airline commuter airline domestic airline feeder airline international airline scheduled airline transport pilot
airspace <i>n</i> advisory airspace [ADA] airspace boundary airspace capacity airspace chart airspace class airspace classification airspace congestion airspace design airspace reservation airspace restriction air traffic services airspace approach control airspace ATS airspace block of airspace classification of airspace controlled airspace [CAS ²] (...) ukupno 37 podnatuknica	airspace airspace denial	airspace	airspace airspace classification airspace control area airspace restrictions airspace vertical division airspace warning area advisory controlled airspace

<p>runway <i>n</i> active runway available runway length downhill runway slope grass runway inactive runway intersecting runway landing runway paved runway precision approach runway runway category runway centre line runway code runway direction runway end lights runway in use runway layout runway marking runway touchdown zone runway visual range [RVR] unpaved runway</p> <p>(...) ukupno 111</p>	<p>runway runway alignment runway alert system runway alignment factor runway alignment indicator runway basic length runway capacity runway condition runway controller runway designator runway direction numbers runway-edge lights runway-end lights, REIL runway-end safety area</p> <p>(...) ukupno 29</p>	<p>runway active runway alignment with the runway clear runway a level runway runway visual range runway in use</p>	<p>runway runway alignment indicator runway arresting gear runway basic length runway centerline runway centreline lighting runway centreline lights runway condition reading runway control van runway controller runway designation marking runway edge lights runway end identifier lights runway end light</p> <p>(...) ukupno 52</p>
---	--	---	--

2. Usporedba obradbe kolokacija u analiziranim zrakoplovnim rječnicima

Automatsko izlučivanje kolokacija

Pojačano zanimanje za izučavanje kolokacije logično je, osobito posljednjih desetak godina, potaknulo iznalaženje djelotvornih načina automatskoga izlučivanja kolokacija. U jeziku struke to razumijeva računalno prepoznavanje pojmovnih odnosno semantičkih veza. Takvo prepoznavanje, ističe Wanner (2007: 128), mora uzeti u obzir, s jedne strane, paradigmatske odnose (sinonimiju, antonimiju i hiperonimiju), a s druge sintagmatske odnose oslanjajući se na tzv. kolokacijsku tipologiju.

Za engleski jezik već postoje brojni razvijeni alati za automatsko izlučivanje kolokacija, no u najnovije vrijeme zapažen se uspjeh u tom smjeru bilježi i u hrvatskom računalnom jezikoslovlju.

Razmjerno tradicionalan pristup primijenjen u izradbi ovoga priručnika, oslonjen na minucioznu analizu korpusa, bilo je naposljetku zanimljivo usporediti s najnaprednijim takvim dosada razvijenim programom u nas (Delač et al., 2009) koji nam je u međuvremenu, zahvaljujući susretljivosti autora, postao dostupnim.

Korpus na kojem se temelji *Aviation English Terms and Collocations* podvrgnut je programu za automatsku ekstrakciju kolokacija s pomoću alata TermeX.³ Analizom korpusa u TermeXu dobili smo 81 439 bigrama, 106 848 trigramama i 70 036 kvadrigrama u frekvencijskim popisima. Za usporedbu je uzeto 1000 po frekvenciji najbolje rangiranih automatski dobivenih kolokacija — 300 bigrama, 500 trigramama i 200 kvadrigrama. Tih je 1000 kolokacija zatim uspoređeno s kolokacijama u rječniku, s tim da pravopisne razlike i razlike među pojavnicama u jednini i množini nisu uzete u obzir (npr. *cross country flight* i *cross-country flight*, *center of gravity* i *centre of gravity*, *call sign* i *call signs* itd.). Zabilježene su sve nove pojavnice, odnosno one kolokacije koje nisu bile navedene u rječniku.

Prvim je pregledom dobiven popis od 534 kolokacije, koji je sljedećim pregledom smanjen na 280. Izbačene su sve kolokacije koje su nastale pogreškom, poput *message element URG*, a potom i sve leksičke sveze koje nisu prave kolokacije (*mode S*). Pod tim razumijevamo i kolokacije koje sadrže veznike *or* i *and* (*day or night*, *field or symbol*, *aircraft and vehicles*, *direction and speed*), kao i kolokacije koje su zapravo nepotpune rečenice: *drag*

³ Najsrdačnije zahvaljujemo Davoru Delaču na pomoći i savjetima tijekom analize rezultata.

increases, controller is responsible, aircraft is flying. Automatskom su ekstrakcijom dobivene i kolokacije u kojima je pridjev u komparativu (*colder air, lower levels*) i mjerne jedinice poput *force per g, per hour* i *per minute*.

U posljednjem su čitanju izbačene sve kolokacije za koje je odlučeno da ne pripadaju isključivo zrakoplovstvu (*meet the requirement, member state*) i one na koje su sastavljači naišli pri sastavljanju rječnika, ali su odlučili uvrstiti njihove frekventnije oblike. Takvi su primjeri *coefficient of drag, category of aircraft, change in pressure*, itd. Umjesto njih naš rječnik navodi *drag coefficient, aircraft category* i *pressure change*. Izbačene su i dvočlane ili tročlane kolokacije kod kojih se provjerom utvrdilo da im nedostaje jedan element ili da su nastale od duljih kolokacija. U primjerima poput *lines of magnetic* i *engine aeroplane* jasno je da se radi o “nepotpunim” kolokacijama i da umjesto njih pravi izrazi glase *lines of magnetic flux* i *single engine aeroplane*. Zadnje je čitanje donijelo konačan popis od 140 kolokacija koje bi se terminološkom obradbom mogle unijeti u specijalizirani rječnik zrakoplovnoga nazivlja.

Analizom 1000 kolokacija dobivenih s pomoću programa za automatsku ekstrakciju kolokacija i njihovom usporedbom s kolokacijama navedenima u *Aviation English Terms and Collocations*, utvrđeno je da bi se ovom metodom dobila 641 kolokacija, od kojih bi 606 bilo uvršteno u rječnik. U našem je popisu više od 500 tih kolokacija već zastupljeno, a ostalih stotinjak bi se nakon terminološke obradbe moglo uzeti u obzir za uvrštavanje. Najveća je prednost ovakva načina rada brže dobivanje građe za daljnju terminološku obradbu. Manji su nedostaci relativno velik broj tročlanih pojava koje se javljaju u frekvencijskim popisima, ali nisu prave kolokacije jer im nedostaje jedan element u leksičkoj svezi. U opsežnijem bi se korpusu taj broj sigurno smanjio i rezultati bi bili bolji.

S obzirom na to da je program kojim smo se služili isto namijenjen “poluautomatskoj” obradbi kolokacija, pružajući korisniku mogućnost da preko sučelja sam označi koja je pojava kolokacija a koja to nije, rezultati dobiveni ovim putem uvelike se podudaraju s rezultatima dobivenima izvlačenjem konkordancija u Wordsmithu. Ono što je, međutim, bitno primijetiti jest da odabir kolokacija iz korpusa, određivanje njihova stvarna oblika i broja članova te njihova precizna semantička interpretacija nije moguća bez temeljita uvida u korpusnu građu, a upravo je to najdugotrajniji i najzahtjevniji dio posla kada je riječ o prikupljanju leksikografskih podataka.

Automatska ekstrakcija kolokacija svakako ubrzava rad, no ljudsko je znanje i dalje bitan čimbenik u izradi terminoloških leksikona i specijalizi-

ranih rječnika. Pritom se pokazuje da pri odabiru relevantne leksičke građe statističke mjere same po sebi nisu pouzdan vodič kao ni frekvencija pojedinih leksičkih supojavljivanja jer čestota pojavljivanja nekoga naziva ili terminologizirane leksičke sveze u korpusu, ma kako reprezentativan on bio, ne korespondira jednoznačno s njezinom važnošću u ukupnosti strukovnog vokabulara.

Ono što se, međutim, pokazalo vrijednom spoznajom u radu na opisanom rječniku jest sposobnost čovjeka (ne nužno izvornoga govornika, a u ovom je konkretnom slučaju bila mahom riječ o studentima ili apsolventima studija aeronautike) da prepozna granice kolokacije, odnosno razlikuje kolokacije od nekolokacija. U pedagoškom kontekstu koji je i potaknuo izradbu ovoga priručnika i koji je u metodološkom smislu do neke mjere moguće ugraditi i u nastavu jezika struke i taj se aspekt čini potencijalno produktivnim.

Literatura

- Bergenholtz, Henning, Sven Tarp (ur.). 1995. *Manual of Specialized Lexicography*. Amsterdam, Philadelphia : John Benjamins Publishing Company. 255 str.
- Delač, Davor, Zoran Krleža, Jan Šnajder, Bojana Dalbelo Bašić, Frane Šarić. TermeX: A Tool for Collocation Extraction. 2009. U knj. *Technologies for the Processing and Retrieval of Semi-Structured Documents. Experience from the CADIAL Project*. Priredili Marko Tadić, Bojana Dalbelo Bašić, Marie-Francine Moens. Zagreb : Croatian Language Technologies Society.
- Greenbaum, Sidney. 1974. Some verb intensifier collocations in American and British English. *American speech* 49, 79–89.
- Heid, Ulrik. 2001. Collocations in Sublanguage Texts: Extraction from Corpora. U knj. *Handbook of Terminology Management*. Vol 2. Priredili Sue Ellen Wright, Gerhard Budin. Amsterdam, Philadelphia : John Benjamins Publishing Company. 788–808.
- Krishnamurthy, Ramesh. 2000. Collocations from *silly ass* to lexical sets. Priredili C. Heffer, H. Sauntson, G. Fox. *Words in Context: A Tribute to John Sinclair on his Retirement*. Birmingham : University of Birmingham. 31–47.
- L’Homme, Marie-Claude, Patrick Leroyer. 2009a. Combining the semantics of collocations with situation-driven search paths in specialized dictionaries. *Terminology* 15:2, 258–283.
- L’Homme, Marie-Claude, Bertrand Claudine. 2009b. Specialized Lexical Combinations: Should they be Described as Collocations or in Terms of

- Selectional Restrictions. <http://www.olst.umontreal.ca/pdf/berlhom.pdf> (20. lipnja 2009.).
- International Civil Aviation Organization. 2004. Manual on the Implementation of ICAO Language Proficiency Requirements. http://caa.gateway.bg/upload/docs/9835_1_ed.pdf (13. studenoga 2009.).
- Sinclair, John et al. (2004) *English Collocation Studies: The OSTI Report (Research in Corpus and Discourse)*. London & New York : Continuum.
- Siepmann, Dirk. 2005. Collocation, Colligation and Encoding Dictionaries. Part I: Lexicological Aspects. *International Journal of Lexicography* 18:4, 409–443.
- Stubbs, Michael. 2001. On inference theories and code theories: corpus evidence for semantic schemas. *Text* 21/3. 437–465.
- Walker, Crayton. 2009. The Treatment of Collocation by Learners' Dictionaries. Collocational Dictionaries and Dictionaries of Business English. *International Journal of Lexicography* 22:3. 3281–3299.
- Wanner, Leo, Bernd Bohnet, Mark Giereth, Vanesa Vidal. 2007. The first steps towards the automatic compilation of specialized collocation dictionaries. U knj. *Application Driven Terminology Engineering*. Priredili Fidelis Ibekwe-SanJuan, Anne Conamines, M. Teresa Cabré Castelyvi. Amsterdam/Philadelphia : John Benjamins Publishing Company. 127–161.

Rječnici

- Bratanić, Maja, Ana Ostroški Anić, Tomislav Radišić (ur.). 2010. *Aviation English Terms and Collocations. An Alphabetical Checklist*. Zagreb : Sveučilište u Zagrebu, Fakultet prometnih znanosti.
- Crocker, David. 2007. *Dictionary of Aviation*. Second Edition. London : A & C Black Publishers Ltd.
- Eleršek, Leonard. 2008. *Englesko-hrvatski zrakoplovni rječnik*. Zagreb : Ministarstvo obrane Republike Hrvatske.
- Gunston, Bill. 2004. *The Cambridge Aerospace Dictionary*. Jane's Information Group. New York : Cambridge University Press.
- Stevenson, Angus (ed.). *Shorter Oxford English Dictionary*, 2007. Oxford : Oxford University Press.

A Pedagogical Approach to a Corpus-Driven Compilation of an LSP Collocational Dictionary

Abstract

This paper describes the compilation of a specialized and typologically innovative terminological collocational dictionary. Based on a targeted corpus of Aviation English, an extensive list of almost 10000 aviation English terms and their typical lexical combinations was compiled with the aim of precisely rendering the LSP terminology and vocabulary globally prescribed for pilots and air traffic controllers. A rather simple concordancing tool was used, thus enabling several senior students of aeronautics to take part in the compilation process. Such a partial computer-assisted approach to terminology extraction demonstrated some advantages not only over traditional lexicographic and terminographic methods but also over more sophisticated computational methods, which do not take into account all criteria for selecting terminologically relevant collocations. The lexical material obtained by this procedure can be used as a potential “left side” of any bilingual aviation dictionary with English as a target language. Moreover, it provides ample potential for further lexicological research and can also be used for the interactive acquisition of (foreign language) terminology.

Ključne riječi: strukovno nazivlje, terminološki kolokacijski rječnik, specijalizirani korpus, izlučivanje kolokacija

Key words: special purpose language terminology, terminological collocational dictionary, specialized corpus, collocation extraction