

Korpus metafor KOMET 1.0

Špela Antloga

Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru
Koroška cesta 46, 2000 Maribor
s.antloga@um.si

1 Uvod

Prepoznavanje vrednosti in razširjenosti metaforičnih izrazov v jeziku je v zadnjih dvajsetih letih vodilo k povečanemu zanimanju za sistematično identifikacijo metaforičnih izrazov in metafor v posameznih jezikih, kljub temu pa je identifikacija metaforično rabljenega jezika še vedno eden od izzivov, s katerim se soočata sodobno korpusno in računalniško jezikoslovje.

Po sodobni definiciji metafore niso samo jezikovni izraz, ampak v metaforah tudi razmišljamo. Med različnimi teoretičnimi pristopi, ki so se konec 70. let prejšnjega stoletja v času t. i. kognitivnega preobrata posvečali preučevanju metafore, je še danes ena najvidnejših zagotovo teorija konceptualne metafore, ki so jo razvili George Lakoff in njegovi sodelavci (Lakoff in Johnson, 1980; Lakoff in Turner, 1989). Po omenjenem teoretičnem modelu so metafore bistven element človekovega spoznavanja in sredstvo, ki nam omogoča, da razumemo in doživljamo eno izkušnjsko področje ali domeno (*domain*) s pomočjo (v okviru) drugega (Bratož 2010: 19). Prenos poteka s t. i. medpodročnimi preslikavami (*cross-domain mappings*) med izhodiščnim področjem (*source domain*), ki je običajno konkretnije, in ciljnim področjem (*target domain*), ki je bolj abstraktno. Jezikovne, vizualne in multimodalne uresničitve konceptualnih metafor pa nam omogočajo vpogled v te medpodročne preslikave.

Na prelomu tisočletja so kritiki kognitivistom očitali subjektivnost, neempiričnost in nekonsistentnost pri prepoznavanju (iskanju) in razlagi metafor. Kot odziv na tovrstne očitke je skupina raziskovalcev¹ za angleščino² razvila postopek za ugotavljanje metaforičnih besed v besedilu, poimenovan MIPVU (Steen in drugi, 2010),³ da bi omogočila objektivnejšo, natančnejšo in bolj sistematično (jezikoslovno) analizo metaforičnih izrazov v različnih besedilih. Temeljno izhodišče za označevanje metaforičnih besed pri tem postopku je ugotavljanje razmerja med osnovnim in kontekstualnim pomenom besede. Pri tem je treba za vsako leksikalno enoto ugotoviti, ali se njen konkretni kontekstualni pomen razlikuje od njenega osnovnega pomena. Eden od poskusov oblikovanja korpusa metafor v slovenščini, ki bi omogočal jezikoslovno analizo metaforičnih izrazov in metafor v različnih besedilih ter ponujal možnost za prepoznavanje kulturnospecifičnega pomena metafor, je korpus metafor KOMET 1.0.⁴

2 Korpus metafor KOMET 1.0

Korpus KOMET (KORpus METafor) 1.0 je korpus slovenskih publicističnih, leposlovnih in spletnih besedil v obsegu 200.622 besed, v katerem so ročno označene metaforične besede, torej besede, katerih pomen v danem kontekstu odstopa od osnovnega pomena te besede. Označevanje metaforičnih besed je temeljilo na omenjenem postopku za identifikacijo metafor MIPVU, kar je pomenilo označevanje vseh jezikovnih elementov, povezanih s pomenskimi prenosi na osnovi medpodročnih preslikav. Označenim metaforičnim besedam je bilo nato pripisanih 69 pomenskih shem, ki funkcionirajo kot sistem kategorij, ki so strukturirane glede na določen kontekst, ki jih motivira. To nam omogoča, da znotraj določene pomenske kategorije (npr. naravni pojavi, čas, prostorska orientacija, družina, premikanje itd.) prepoznamo metaforične besede, ki so lahko potencialno uresničitev neke konceptualne strukture.

¹ Skupina se je poimenovala Pragglejaz po prvih črkah imen raziskovalcev, ki so sodelovali projektu: Peter Crisp, Ray Gibbs, Alan Cienki, Graham Low, Gerard Steen, Lynne Cameron, Elena Semino, Joe Grady, Alice Deignan in Zoltan Kövecses.

² Postopek je s prilagoditvami značilnostim posameznih jezikov sprožil zanimanje za identifikacijo metaforičnih izrazov in metafor v češčini, litovščini, madžarščini, poljščini, srbsščini, španščini ter za izdelavo korpusov metafor v angleščini (Steen G. J. idr., 2010), ruščini (Yulia Badryzlova in Olga Lyashevskaya, 2017) in hrvaščini (Kristina Despot idr., 2019).

³ Metaphor Identification Procedure Vrije Universitāt.

⁴ Projekt izdelave korpusa je bil financiran v okviru projekta CLARIN.si 2019. Korpus je dostopen na naslovu <http://hdl.handle.net/11356/1293>.

2.1 Metodološka izhodišča za izdelavo korpusa

Osnovna baza je bila oblikovana z uravnoteženim naborom zvrstno različnih besedil iz korpusa MAKS.⁵ Za označevanje metaforičnih besed in pomenskih shem je bilo uporabljeno orodje WebAnno, kjer je postopek identifikacije metaforičnih besed potekal po naslednjih korakih:

1. Branje celotnega besedila za splošno razumevanje pomena.
2. Za vsako leksikalno enoto v besedilu je bil določen njen pomen v kontekstu. Upoštevano je bilo okolje, v katerem se leksikalna enota nahaja (kaj je pred in za njo).
3. Za vsako leksikalno enoto je bilo ugotovljeno, ali ima osnovnejši pomen v drugih kontekstih. Osnovni pomeni so običajno:
 - konkretnější (to pomeni, da si je tisto, kar evocirajo, lažje predstavljati, videti, slišati občutiti, vohati ali okusiti),
 - povezani s telesnim delovanjem,
 - bolj natančni in jasni.
4. Če se je kontekstualni pomen razlikoval od osnovnega, je bila beseda označena kot metaforična beseda (MRW).

V naslednjem koraku je bila označenim metaforičnim besedam dodana tudi oznaka 'frame' za določanje pomenske sheme. S tem je bila označena beseda uvrščena v določeno pomensko področje, ki je bilo opredeljeno glede na osnovni pomen označene besede in funkcionira kot »izhodiščna domena« v pojmovanju konceptualne metafore.

2.1.1 Postopek določanja metaforičnih besed (oznaka 'MRW')

Pri postopku identifikacije so bile označene leksikalne enote (glagoli, samostalniki, pridevniki, prislovi in predlogi oz. stalne besedne zveze), katerih kontekstualni pomen se je razlikoval od njihovega osnovnega pomena. Osnovni pomen je bil določen na podlagi hierarhije slovarskih pomenov danega leksema.⁶ Označenim metaforičnim besedam je bila nato pripisana informacija o tem, ali gre za (1) indirektno metaforo, (2) direktno metaforo ali (3) mejni primer. Označeni so bili tudi (4) metaforični signalizatorji.

(1) Z oznako 'indirektna metafora' (MRWi) so bili označeni kontekstualni pomeni besed, ki niso osnovni pomeni teh besed in jih lahko pojasnimo kot preslikavo z ene domene na drugo. Npr. v povedi

Lepi, slavni in bogati naravnost obožujejo knežjo družino, ki vsakič poskrbi za pikantne govorice

sta bili z oznako MRWi označeni besedi *naravnost* in *pikantne*. Osnovni pomen besede *naravnost* namreč izraža gibanje ali položaj, osnovni pomen besede *pikanten* pa prekomerno začinjenost. Kontekstualni pomen je sicer že lahko konvencionaliziran in zato opredeljen v slovarju, npr. pomen intenzivnosti (tretji pomen iztočnice *naravnost*) ali privlačnosti (drugi pomen iztočnice *pikanten*). Vendar to ne pomeni, da je že vsaka pomenska enota z več pomeni označena kot metaforična, če v konkretnem primeru ni rabljena v slovarskem prvem pomenu. Kontekstualni pomen mora biti dovolj oddaljen od osnovnega pomena, da je označen kot metaforičen.

(2) Z oznako 'direktna metafora' (MRWd) so bile označene besede, ki so sicer rabljene v svojem osnovnem pomenu, a je kontekstualni pomen lahko pojasnjen kot preslikava z ene domene na drugo. Npr. v povedi

Obnašate se kot sestradani golob, ki kljuva kot zmešan, prepričan, da je samo za las oddaljen od cilja

sta besedi *sestradan* in *golob* sicer rabljeni v svojem osnovnem pomenu 'ki je zaradi pomanjkanja hrane zelo izčrpan, oslabel' in 'domača ali divja ptica z majhno glavo, kratkim vratom in krepkim telesom', vendar sta

⁵ Korpus MAKS v obsegu približno 10 milijonov besed z besedili iz publicistike, leposlovja in spletnih portalov je bil izdelan v okviru projekta Slovenščina na dlani. Več na <http://projekt.slo-na-dlani.si/sl/>.

⁶ V uvodu v SSKJ (§ 75) je navedeno, da je na prvem mestu večpomenskih iztočnic osnovni pomen, to je tisti, ki je v sodobnem knjižnem jeziku najbolj nevtralen oziroma prevladujoč. Nadaljnji pomeni so razporejeni po stopnji odvisnosti od osnovnega pomena. Včasih na prvem mestu ni najmočnejši, ampak pomensko izhodiščni pomen iztočnice, zlasti če je to potrebno zaradi razvrstitve nadaljnjih pomenov (§ 76). Kadar ima beseda več pomenov, ki so enako močni, stoji na prvem mestu pomen, ki ima največ ali najmanj pomenskih prvin. Nadaljnji pomeni si sledijo, kakor število teh elementov ali upada ali narašča (§ 77). Manjši pomenski premiki so obdelani kot podpomeni ali pomenski odtenki v okviru nadrejenega pomena (§ 78). Pri slovnicih besedah odloča o razvrščanju razlag pogostnost, ob enaki pogostnosti pa ustaljeno slovnico zaporedje (§ 79).

skupaj s primerjalnim veznikom *kot* (metaforičnim signalizatorjem) zaradi preslikave pomena z izhodiščne živalske domene na človeka označeni kot direktno rabljeni metaforični besedi. Enotam, ki so bile označene kot MRWd, ni bila dodana oznaka 'frame' za pomensko shemo, saj je ta že neposredno izražena.

(3) Kot mejni primeri (WIDL) so bile besede označene takrat, ko iz konteksta ni bilo mogoče določiti, ali gre za metaforični ali osnovni pomen besede. Npr. v izseku intervjuja, kjer nogometni igralec na vprašanje odgovori: »Do derbija se je že nakopičila utrujenost ob nizu zaporednih tekem,« novinar doda, da se je *mladi hrvaški reprezentant tako kot večina najprej dotaknil bolečih mišic*. V tem primeru ni mogoče določiti, ali gre za osnovni pomen glagola *dotakniti se* ('približati se tako, da pride do dotika, stika') ali metaforični ('na kratko, nekoliko spregovoriti o čem'), zato je enota označena kot mejni primer.

(4) Z oznako metaforični signalizatorji (MFlag) so bile označene besede, ki signalizirajo posamezne besede in slovnične strukture, npr. *kot, kakor, enačiti* (kaj) *z, metaforično (rečeno)* ipd.

Posebej so v korpusu označene še frazeološke enote (ki so razumljene kot ena leksikalna enota z enim pomenom, ki pa je metaforičen), metonimije (pri čemer ne gre za preslikavo pomena z ene domeno na drugo, ampak se preslikava dogaja znotraj ene (iste) domene) in poosebitve.

2.1.2 Postopek določanja pomenskih shem (oznaka 'frame')

Z oznako 'frame' je bila predhodno označena metaforična beseda uvrščena v določeno pomensko shemo, ki je bila opredeljena glede na osnovni pomen označene besede (pomen, iz katerega izhaja preslikava). Določanje pomenske sheme označenim metaforičnim besedam omogoča, da znotraj določene pomenske kategorije (npr. čas, prostorska orientacija, premikanje, čustvovanje itd.) prepoznamo metaforične izraze, ki so lahko potencialno uresničitev določene konceptualne strukture. Trenutno je označenim metaforičnim besedam dodanih 69 pomenskih shem. Seznam pomenskih shem ni končen in se lahko ob dodajanju novih besedil še dopolnjuje. Npr. v povedi

Zato je telekomunikacijska družba pripravila napad na oba monopolista na evropskem trgu

je bila beseda *napad* zaradi kontekstualnega pomena označena kot indirektna metaforična beseda. Pripisana ji je bila pomenska shema 'boj', kar izhaja iz osnovnega pomena označene besede 'oborožena vojaška akcija z namenom uničiti, onesposobiti nasprotnikove sile'.

2.1.3 Splošna vodila pri označevanju metaforičnih besed glede na besedno vrsto

Za označevanje metaforičnih besed so bila določena splošna vodila za posamezne besedne vrste. Osnovna enota za označevanje je leksikalna enota, ki je opredeljena kot beseda, stalna besedna zveza, zveza glagola in členka/predloga ali frazem z enim samostojnim pomenom.

Kontekstualni pomen **glagolov** je bil določen glede na hierarhijo pomena posameznega glagola v SSKJ, pri čemer kontekstualni pomen, ki sicer v slovarju ni označen kot osnovni pomen tega glagola, a vsebuje kvalifikator »s širokim pomenskim obsegom«, ni označen kot metaforična beseda.

Kontekstualni pomen **samostalnikov** in **pridevnikov** je bil določen glede na hierarhijo pomena posamezne besede v SSKJ.

Večina označenih **prislovov** je zaenkrat dobila oznako 'WIDL' (mejni primeri), saj slovar načeloma eksplicitno ne hierarhizira pomenov prislova, ampak skupaj s primeri navaja zgolj glagol ali pridevnik, iz katerega je tvorjen. Zgolj na podlagi tega osnovnega pomena posameznega prislova ni mogoče določiti, saj ne predpostavljam, da se hierarhija pomenov ohranja pri prehodu v drugo besedno vrsto (enako velja za glagolnike, če nimajo določenih pomenov).

Določanje kontekstualnega pomena **predlogov** izhaja iz tega, da primarno poimenujejo prostorsko razmerje, čeprav je pri hierarhiji pomenov v slovarju upoštevana pogostnost (gl. opombo 6). Za razliko od označevalne sheme za angleščino je za vsak predlog osnovni pomen določen še po sklonu, s katerim se veže, npr. *pod* + tožilnik (za izražanje premikanja, usmerjenosti k spodnji strani česa) ali npr. v metaforičnem pomenu 'biti *pod* vplivom koga' oziroma *pod* + orodnik (za izražanje položaja na spodnji strani česa) ali npr. v metaforičnem pomenu 'zlomiti se *pod* pritiskom'.

2.2 KOMET 1.0 v številkah

V korpus metafor KOMET 1.0 v obsegu 200.622 besed je vključenih 42 publicističnih besedil v obsegu 95.531 besed (48 % celotnega korpusa), 8 strokovnih besedil v obsegu 23.732 besed (12 % celotnega korpusa), 5 leposlovnih besedil v obsegu 38.030 besed (19 % celotnega korpusa) in spletna besedila v obsegu 43.329

besed (21 % celotnega korpusa). Od 10.632 označenih metaforičnih besed (5,3 % vseh besed) je bilo največ implicitnih metafor (MRWi), in sicer 86,3 %, kot mejnih (WIDL) je bilo označenih 11,6 % vseh označenih metaforičnih besed, najmanj je označenih direktnih metafor (MRWd), in sicer 2,1 %. Besednovrstno je 39,2 % vseh označenih metaforičnih besed glagolov, 27,6 % samostalnikov, 13,1 % pridevnikov, 12,7 % predlogov in 7,4 % prislovov. Med pripisanimi 69 pomenskimi shemami prevladujejo (po vrsti): premikanje, vidno zaznavanje, telo, razsežnost, življenjski procesi, čas, boj, smer, oblika, tekočina, gib, potovanje, ogenj, objekti, lov itd.

3 Zaključek

Za vpogled v metaforično konceptualizacijo stvarnosti je zaradi kulturne razsežnosti metafore smiselno združevanje različnih teoretičnih pristopov. Tradicionalni kognitivni vidik preučevanja konceptualnih struktur smiselno nadgrajuje korpusni pristop, ki omogoča natančnejšo in bolj sistematično analizo metaforičnih izrazov v različnih besedilih. Namen izdelave korpusa metafor ni bilo iskanje niti interpretacija metafor. Postopek identifikacije metafor je namreč temeljil na predpostavki, da ima vsak pomen besede, ki v določenem kontekstu odstopa od osnovnega pomena te besede, potencial, da predstavlja jezikovno uresničitev neke (konceptualne) strukture. Za nadaljnje delo s korpusom bodo še potrebne izboljšave, predvsem pri prepoznavanju metaforičnosti v povezavi s posameznimi skloni v smislu »osnovnega pomena«, ki ga izraža posamezen sklon, pri opazovanju metaforičnosti zvez glagola s členki in predlogi ter pri natančnejši opredelitvi pomenskih shem. Že sedaj pa korpus omogoča kvantitativno analizo, primerjavo metaforično rabljenih besed in pomenskih shem glede na besedilni tip/žanr, analizo izrazov znotraj določenega pomenske sheme, analizo posameznih izrazov, ki so potencialno uresničitev več konceptualnih struktur ipd.

4 Literatura

- Špela Antloga. 2010. Metaphor corpus KOMET 1.0, Slovenian language resource repository CLARIN.SI. <http://hdl.handle.net/11356/1293>.
- Yulia Badryzlova in Olga Lyashevskaya. 2017. Metaphor Shifts in Constructions: the Russian Metaphor Corpus. V: *Computational construction grammar and natural language understanding: Papers from the 2017 AAI Spring Symposium*. The AAI Press.
- Silva Bratož. 2010. *Metafore našega časa*. Fakulteta za management, Koper.
- Lyanne Cameron. 1999. Operationalising 'metaphor' for applied linguistic research. V: *Researching and applying metaphor*, str. 3–28. Cambridge University Press.
- Richard Eckart de Castilho, Éva Mújdricza-Maydt, Seid Muhie Yimam, Silvana Hartmann, Iryna Gurevych, Anette Frank in Chris Biemann 2016. A Web-based Tool for the Integrated Annotation of Semantic and Syntactic Structures. V: *Proceedings of the LT4DH workshop at COLING 2016*. <https://webanno.github.io/webanno/>.
- Alan Cienki. 2008. Why study metaphor and gesture. V: *Gesture Studies*, str. 5–25. John Benjamins, Amsterdam/Philadelphia.
- Kristina Despot, Mirjana Tonković, Mario Brdar, Benedikt Perak, Ana Ostroški Anić, Bruno Nahod, Ivan Pandžić. 2019. MetaNet.HR: Croatian Metaphor Repository. V: *Metaphor and Metonymy in the Digital Age. Theory and Methods for Building Repositories of Figurative Language*, str. 123–146. John Benjamins Publishing Company.
- Raymond W. Gibbs. 1999. Researching Metaphor. V: *Researching and applying metaphor*, str. 29–47. Cambridge University Press.
- George Lakoff in Mark Johnson. 1980. *Metaphors We Live By*. University of Chicago Press.
- George Lakoff in Mark Turner. 1989. *More than Cool Reason: A Field Guide to Poetic Metaphor*. The University of Chicago Press.
- Graham Low. 1999. Validating metaphor research projects. V: *Researching and applying metaphor*, str. 48–68. Cambridge University Press.
- Pragglejaz Group. 2007. MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol* 22 (1): 1–39.
- Elena Semino. 2017. Corpus linguistics and metaphor. V: *The Cambridge Handbook of Cognitive Linguistics*, str. 463–476. Cambridge University Press.
- Slovar slovenskega knjižnega jezika, druga, dopolnjena in deloma prenovljena izdaja*. www.fran.si.
- Gerard J. Steen, Aletta G. Dorst, Berenike J. Herrmann, Anna A. Kall, Tina Krennmayr in Tryntje Pasma. 2010. *A method for linguistic metaphor identification. From MIP to MIPVU*. John Benjamins, Amsterdam/Philadelphia.