

# Formulaic Sequence Annotation Guidelines, Croatian supplement

Jan Šnajder

September 19, 2016

Upute za označavanje za engleski jezik u vrijede i za hrvatski jezik. Primjeri formulaičnih izraza u hrvatskome (popis nije iscrpan):

- Imeničke fraze:
  - “električna vuča”, “kontrola kvalitete”, “investicijski fond”, “drugi svjetski rat”, “Katica za sve”, “mali od palube”
  - Tu su osobito česte kombinacije pridjev+imenica ili imenica+imenica-u-genitivu, a moguće su naravno i složenije kombinacije
- Imena poznatih entiteta (osoba, mjesta, organizacija) – sintaktički gledano, to su također imeničke fraze:
  - “Zagrebačka županija”, “Fakultet elektrotehnike i računarstva”, “Justin Bieber”, “Marks and Spencer”, “Balkanska ruta”
- Povratni glagoli:
  - Pravi povratni (riječ “se” korištena je kao zamjenica i može se zamijeniti sa “sebe”): “predati se”, “upustiti se”
  - Uzajamno povratni: “grliti se”, “pozdraviti se”,
  - Nepravi povratni (nema prave povratnosti radnje na subjekt, “se” je čestica i ne može se zamijeniti sa “sebe”): “glasati se”, “veseliti se”, “čuditi se”, “smijati se”
- Glagolske fraze:

- “pridobiti pažnju”, “biti u žiži”, “dati sve od sebe”, “baciti \* u vjetar”, “može se govoriti o \*”, “dolaziti na naplatu”, “stići na naplatu”
- Pridjevske fraze:
  - “koliko toliko”, “s time povezan”, “bez presedana”
- Priložne oznake:
  - “bezbroy puta”, “svakim danom sve \*”, “malo puno”
- Prijedložne oznake:
  - “u bližoj okolini”, “u nepoznatom smjeru”, “u ovom trenutku”
- Diskursni konektori:
  - “sve dok je”, “shodno tome”, “drugim riječima”, “na primjer”
- Klišeizirane fraze:
  - “plan i program”, “može \* biti sram”, “mama ti mogu biti”, “čiča miča gotova je priča”, “bolje išta nego ništa”, “nepoznati netko”, “možda su u šumi”, “nema govora o \*”

U odnosu na engleski, kod hrvatskoga jezika nailazimo na neke specifičnosti, ponajviše zbog razlika u morfologiji (visok stupanj fleksije i derivacije) i sintaksi (relativno slobodan poredak riječi u rečenici). S druge strane, u hrvatskome ne postoje neki fenomeni koji postoje u engleskome, npr., članovi i frazalni glagoli.

## Morfologija

Za razliku od engleskog, hrvatski je visoko flektivan jezik, što znači da se jedna te ista fraza može pojavljivati u različitim flektivnim oblicima, npr., “jedan te isti”, “jedna te ista”, “jednog te istog” (primijetite, međutim, da je u nekim slučajevima fraza okamenjena i ne dopušta nikakvu morfološku varijaciju, npr., “Katicama za sve” ili “u bližim okolicama” nisu formulaički nizovi). Pri označavanju, fraza će biti prikazana u obliku u kojem

se ona najčešće nalazi u korpusu; taj oblik ne mora nužno biti nautknički oblik kakav biste našli u rječniku. Npr., formulaički niz “dvostruka igra” pojavljuje se u korpusu najčešće u kosom obliku “dvostruku igru” (jer se fraza najčešće koristi kao objekt u akuzativu), pa će taj oblik biti prikazan, međutim oba oblika se odnose na isti formulaički izraz i trebaju se tretirati ravnopravno. Dodatno će biti prikazan i potpuno lematizirani oblik izraza (“dvostruki igra”), koji može, ali i ne mora, biti gramatički ispravan oblik izraza u hrvatskom jeziku. Pri označavanju treba dakle zanemariti sve ove razlike i označiti formulaičnost niza neovisno o tome u kojem je morfološkom obliku prikazan. Osobito treba obratiti pažnju na supletivne oblike, posebice glagola biti (sam/si/je/smo/ste/su) i slično. Npr., “je u žiži” je formulaični niz, kao što je to i niz “su u žiži” i “smo u žiži”, što su sve flektivne varijante niza “biti u žiži”. Budući da se provodi lematizacija, ovi se nizovi tretiraju kao istovjetni, što znači da se u popisu ne bi smjele zasebno pojaviti varijante ovog niza.

Dodatna razlika hrvatskoga u odnosu na engleski jezik u pogledu morfologije jest ta što hrvatski koristi derivacijsku morfologiju kako bi kodirao neke semantičke značajke poput spola osobe koja obavlja zanimanje (tzv. *mocijski parovi*) ili glagolskog aspekta (svršenost i nesvršenost glagola), npr., “kuhar” – “kuharica”, “dobitnik” – “dobitnica”, “stići” – “stizati”, “komentirati” – “prokomentirati”. Ove razlike neće biti uklonjene lematizacijom, što znači da se, primjerice, fraze “goli kuhar” i “gola kuharica” tretiraju kao različite, te formulaičnost svake od njih treba ocijeniti zasebno.

Druga značajna razlika hrvatskoga u odnosu na engleski jezik jest ta što hrvatski dopušta relativno slobodan poredak riječi u rečenici. Posljedica toga jest da se u nekim formulaičnim nizovima može mijenjati poredak riječi u rečenici (doduše, ta je pojava rijetka, jer formulaičnost obično diktira poredak riječi) ili, što se događa znatno češće, da se neke riječi mogu umetati u formulaične izraze, da bi se zadovoljila sintaktička pravila hrvatskog ili iz stilskih razloga.

## Sintaksa

Za engleski je već bilo objašnjeno da je moguće umetanje riječi u “rupe” (engl. *gaps*) formulaičnog niza. U hrvatskom se relativno često pojavljuju rupe koje mogu popuniti modifikatori. Npr. “stvoriti \* atmosferu” ili “kazna \* zatvora”. Te rupe su često opcionalne, u smislu da je i niz bez rupe također

formulaičan.

Također su moguće rupe koje popunjavaju objekti. Npr., u formulaičnom nizu “baciti \* u vjetar”, rupu će tipično popuniti objekt, izrečen kao imenička fraza ili zamjenica (“baciti novce u vjetar”, “bacio ih u vjetar”, “bacio sve svoje novce u vjetar”). Kao i kod engleskog, u ovakvim slučajevima fraze “baciti novce u vjetar” ili “bacio ih u vjetar” nisu formulaične, jer riječi “novce” i “ih” varijabilan dio niza te ih se može tretirati kao argumente formulaičnog niza. Shodno tome, poopćeni niz “baciti \* u vjetar” jest formulaičan niz (s rupom).

Međutim, za razliku od engleskog, kod hrvatskog će rupe često nastajati zbog slobodnijeg poretka riječi u rečenici, i takve će rupe uvijek biti opcionalne. Npr., formulaični niz “zadnji trenutak” može se pojaviti u verziji s rupom, “zadnji \* trenutak”, gdje rupu može popuniti glagolska kopula (“je”, “smo”, “ste”, “su”, ...) ili neki složeniji niz (npr., “ga je”, “smo ju”, ...). Oba ova niza tretiramo kao formulaična.

Također vrijedi i obrat: formulaični nizovi koji se sastoje od rupe, npr. “baciti \* u vjetar”, mogu se pojaviti bez rupe. Npr., objekt se može naći prije ili poslije formulaičnog niza, kao u “Baca u vjetar dva milijuna kuna.” ili “Ta odluka baca u vjetar sve što je momčad napravila.” Također, opet je moguće da rupu popune riječi koje se tu moraju naći zbog sintakse, npr. glagolska kopula “biti” za tvorbu perfekta ili futura: “bacila je u vjetar”, “bacit će u vjetar”. Također su moguće i kombinacije, gdje rupu zajedno popunjavaju i kopula i objekt, npr. “bacila je sve svoje novce u vjetar”.

U nastavku slijedi sistematizacija fenomena uzrokovanih slobodnim poretkom riječi u rečenici:

- Transpozicija riječi u formulaičnom nizu:
  - Transpozicija zamjeničke enklitike (povratne zamjenice “se”):
    - \* “veseliti se”  $\Rightarrow$  “se veseli”
    - \* “dati si vremena”  $\Rightarrow$  “si dati vremena”
  - Transpozicija glagolske enklitike (kopule):
    - \* “lako je tako”  $\Rightarrow$  “je lako tako”, “tako je lako”
  - U ovakvim slučajevima, sve transponirane varijante formulaičnog niza tretiramo kao istovrijedne. Drugim riječima, dok god su u nizu iste riječi, može ih se različito poredati i opet dobiti formulaični niz (naravno, samo za one poretke riječi koje su u jeziku dopuštene; npr. “dati vremena si” nije formulaičan niz).

- Umetanje riječi u formulaičan niz:
  - Umetanje glagolske enklitike (kopule):
    - \* “zadnji trenutak”  $\Rightarrow$  “zadnji je trenutak”  $\Rightarrow$  “zadnji \* trenutak”
    - \* “hrvatska banka”  $\Rightarrow$  “hrvatska je banka”  $\Rightarrow$  “hrvatska \* banka”
    - \* “zatvoriti poglavlje”  $\Rightarrow$  “zatvorio je poglavlje”  $\Rightarrow$  “zatvoriti \* poglavlje”
  - Umetanje zamjeničke enklitike:
    - \* “zadnji trenutak”  $\Rightarrow$  “zadnji se trenutak”  $\Rightarrow$  “zadnji \* trenutak”
  - Kombinacije:
    - \* “zadnji trenutak”  $\Rightarrow$  “zadnji su ga trenutak”  $\Rightarrow$  “zadnji \* trenutak”
  - Kod umetanja riječi, formulaičan niz s umetnutom riječi ne tretira se kao formulaičan, jer je umetnuta riječ donekle varijabilna (u svim gornjim primjerima na mjesto rupe moguće umetnuti jednu riječ ili kombinaciju više riječi). Drugim riječima, umetnuti dio treba zamijeniti rupom, a da bi se niz tretirao kao formulaičan.

Na primjer, “zatvoriti poglavlje” je formulaičan niz (po kriteriju semantičke neprozirnosti). Zbog slobodnog poretka riječi u rečenici, niz “zatvoriti \* poglavlje” također tretiramo kao formulaičan. Rupu može popuniti kopula (“zatvorila je poglavlje”) ili neka složenija kombinacija (“zatvorili su jučer nakon dugih pregovora poglavlje”), međutim nizovi s popunjenom rupom nisu formulaični jer su previše specifični. Niz “je zatvorila poglavlje” također nije formulaičan, jer kopula “je” nije dio niza.