

Razvrščanje formulaičnih besednih nizov v slovenščini

V6, 12-12-2019

KAZALO

Kratek opis naloge.....	2
Smernice označevanja	3
1. Skladenjska zgradba.....	3
1.1. Strukturno zaključeni nizi.....	3
1.2. Strukturno nezaključeni nizi.....	4
Dvoumnost pri določanju skladske zgradbe	5
2. Pragmatska funkcija.....	6
2.1. Nizi za opisovanje predmetnosti.....	6
2.2. Nizi za vrednotenje	6
2.3. Nizi za upravljanje diskurza.....	7
Dvoumnost pri določanju pragmatske funkcije.....	7
3. Slovaska relevantnost.....	8
3.1. Slovasko relevantni nizi	8
3.2. Slovasko nerelevantni nizi	9
Dvoumnost pri določanju slovasko relevantnosti	10
Potek dela	11

KRATEK OPIS NALOGE

Z računalniškimi metodami smo iz korpusa pisne slovenščine Gigafida in korpusa govornjene slovenščine Gos izluščili seznam besednih nizov, ki se v rabi pojavljajo zelo pogosto, različne statistične mere pa opozarjajo tudi na posebno povezanost med besedami, ki te nize sestavljajo. Cilj naloge je ročno pregledati izluščene nize in jih v skladu s spodnjimi smernicami razvrstiti glede na različne lastnosti – skladijsko zgradbo, pragmatično funkcijo in slovarsko relevantnost.

Primer razvrstitve nekaj najpogostejših nizov v korpusu Gos:

Niz	Skladijska zgradba	Pragmatična funkcija	Slovarska relevantnost
<i>ne vem</i>	zaključeno	vrednotenje	da
<i>ja ja ja</i>	zaključeno	diskurz	ne
<i>v bistvu</i>	zaključeno	vrednotenje	da
<i>dobro jutro</i>	zaključeno	diskurz	da
<i>se mi zdi da</i>	nezaključeno	vrednotenje	ne
<i>je bil</i>	nezaključeno	predmetnost	ne
<i>dame in gospodje</i>	zaključeno	diskurz	da

SMERNICE OZNAČEVANJA

1. SKLADENJSKA ZGRADBA

Z vidika skladske zgradbe nize razvrščamo na (i) **strukturno zaključene** in (ii) **strukturno nezaključene** nize.

1.1. Strukturno zaključeni nizi

Strukturno zaključeni so skladske celovite strukture, ki jim je mogoče pripisati samostojno skladsko vlogo v stavku, besedilu ali besedni zvezi.

Nekaj primerov strukturno zaključenih nizov različnih tipov:

- celotni stavki ali izjave:
 - ***to je res, dobro jutro, a veš, v redu, dame in gospodje, no vidiš, ja ne vem, daj no ...***
 - vključno z opornimi signali tipa ***ja ja ja, no no, ja mhm***

- zveze z vlogo stavčnih členov ali njihovih jeder:
 - osebkali ali predmeti: ***Tina Maze, glava družine, nacionalni interes, petdeset evrov ...***
 - prislovna določila: ***v tem primeru, leta dva tisoč, še nikoli, v Evropski uniji ...***
 - povedki: ***boš videl, nisem vedela, ne bo šlo***; vključno z:
 - povedki s povedkovimi določili (***prepričan sem, ni problem***)
 - kombinacijami glagola *biti* in modalnih prislovov (***je treba, ni mogoče***)
 - kombinacijami povedkov in zaimkov *si/se* (***se dogaja, se spomnim***)
 - zvezami modalnih glagolov in nedoločnikov (***moram povedati***)

- zveze z vlogo stavčnočlenskih prilastkov:
 - ***sto petdeset, bolj ali manj, čim bolj, kar precej, še posebej***
 - ***iz prejšnjega odstavka, po vrsti, iz osemdesetih ...***
 - ***za socialno delo, za okolje, za razvoj***
 - ***ali pa kaj takega, in tako naprej, in pol***

- zveze z vlogo medstavčnih veznikov oz. besedilnih povezovalcev:
 - ***zaradi tega ker, zato da, sploh če, ali pa ...***, vključno z:
 - prislovno modificiranimi vezniki (***takoj ko, šele ko, še posebej če***)
 - deli dvodelnih veznikov (***ne samo, ampak tudi***)
 - ustaljenimi kombinacijami prislovov/členkov in besede *pa* (***sicer pa***)
 - ***tako da, ne glede na to, to se pravi, kot rečeno ...***

Kot strukturno zaključene označimo vse strukturno zaključene dele prilastke ali jedra ne glede na to, ali se v rabi pojavljajo večinoma kot del daljših besednih zvez, npr.:

- jedra brez levih ali desnih prilastkov (tri ***tisoč evrov***, pet ***milijonov dolarjev***)
- prilastki brez jeder (***malo bolj*** kakšen, ***vse sorte*** česa, leta ***dva tisoč*** pet, francoska ***tiskovna agencija*** afp, ministrstvo ***za obrambo***)
- samostalniške zveze brez predlogov (v ***javnem sektorju***, v ***Windsor Parku***, na ***ta način***)

- Kot strukturno zaključene označujemo tudi **vse predložne zveze** (predlog + samostalnik) ne glede na prisotnost morebitnih pomensko obveznih prilastkov:
 - **z naslovom** __, **med letoma** __ in __, **v občini** __
 - **na osnovi** nečesa, **po besedah** nekoga, **za pripravo** nečesa, **v prid** nečemu
 - **v primerjavi** z nečim, **v sodelovanju** z nekom, **za razliko** od nečesa

1.2. Strukturno nezaključeni nizi

Med strukturno nezaključene nize umeščamo vse ostale nize, pri katerih težko govorimo o kakršnikoli skladenjski ali pomenski celovitosti.

Običajno gre za nezaključene dele daljših skladenjsko oz. pomensko zaključenih enot, kot so:

- fragmenti stavkov
 - **da bi se, to pa je, je ja, jaz sem, ne bo nič, se mi zdi da ...**
- fragmenti povedkov, kot so:
 - vezni glagoli brez določila (**se mi zdi** kakšen, **je bil** kaj/kakšen),
 - pomožni glagoli brez glavnega glagola (**ne bomo, naj bi**),
 - modalni glagoli brez nedoločnika (**ne morem, ne smeš**)
- kot fragmente besednih zvez označujemo samo nize, ki jih ni mogoče interpretirati kot strukturno zaključene sestavne dele zvez (jedra ali prilastki, glej zgoraj), zlasti:
 - nezaključene dele predložnih zvez (**za vse** __, **v zadnjih dveh** letih, **ob 17.** uri)
 - nezaključene zveze s predložno vezljivostjo, ki jih prepoznamo po končnem predlogu, kot so:
 - zveze glagolov s predlogi (**govorimo o, velja za, gre za**)
 - zveze prislovov s predlogi (**vključno z, vse do, že od, skupaj z, tik pred**)
 - zveze pridevnikov in zaimkov s predlogi (**odvisna od, ena od, eden izmed**)
 - zveze samostalnikov s predlogi (**možnosti za, pravica do, ministrstvo za, v noči na, odnos do, hvala lepa za**)

Kot strukturno nezaključene označujemo tudi ustaljene zveze tega tipa (»večbesedne predloge«), kot so **v zvezi z, v skladu z, ne glede na**. Tudi zveze tipa **tako kot, več kot, že več kot**.

Med strukturno nezaključene nize prav tako umeščamo pogoste kombinacije funkcijskih besed oz. struktur brez skupne besedilne, stavčne ali besednozvezne funkcije, npr.:

- kombinacije členkov, medmetov, veznikov, ki so del daljše izjave in ne funkcionirajo kot zaključena celota (so razstavljive zveze, prim. *no no* zgoraj): **ja tudi, ja tako da, ja ma, ne že, ali da**
- vsi nizi, ki vključujejo mašilo *eee*: **eee in eee, eee torej, jaz eee**
- ponavljanja kot posledica načrtovanja nadaljevanja govora: **kaj kaj, to to to, za za** (fragmenti daljših enot). Kadar gre za ponavljanje strukturno zaključenih enot, npr. z namenom poudarjanja (*super super, yo yo, glej glej*), niz označimo kot strukturno zaključen.

Dvoumnost pri določanju skladske zgradbe

V splošnem velja, da nizom vedno pripišemo tisto strukturo, ki je glede na analize primerov rabe v korpusu najpogostejša. Če se niz v rabi pojavlja kot zaključena in nezaključena struktura, pripišemo interpretacijo, ki ustreza njegovi najpogostejši vlogi v korpusu.

[nezaključeno] <u>Mislim, da se me nekaj loteva.</u>	[zaključeno] <i>Pride jutri? Mislim, da.</i>
[nezaključeno] <u>Se mi zdi neizrazito.</u>	[zaključeno] <i>To smo že, se mi zdi.</i>
[nezaključeno] <u>Hočem tako, da bo zdržala.</u>	[zaključeno] <i>Me ni doma, <u>tako da pridi jutri.</u></i>

2. PRAGMATIČNA FUNKCIJA

Z vidika pragmatične funkcije nize razvrščamo na (i) nize za **opisovanje predmetnosti**, (ii) nize za **vrednotenje** in (iii) nize za **upravljanje diskurza**. Pragmatično funkcijo pripisujemo vsem nizom, ne glede na strukturno zaključenost niza.

2.1. Nizi za opisovanje predmetnosti

Nizi za opisovanje predmetnosti (angl. *referential expressions*) so rabljeni za poimenovanje konkretnih ali abstraktnih predmetov, bitij, stvari in dogodkov ali njihovih lastnosti. Z njimi govorci oblikujejo jedrno vsebino sporočila, ki ga želijo posredovati naslovniku. Nekaj primerov:

- poimenovanje: ***Evropska unija, nacionalni interes, nič osem nič, d.o.o. ...***
- poročanje in poizvedovanje: ***jaz sem, je bilo, da bi se, to je, kaj je bilo, pol je pa ...***
- primerjava lastnosti: ***v skladu z, v zvezi z, v primerjavi z, v katerem je ...***
- naštevanje: ***in tako naprej, ali pa kaj ...***

[Posebnost] Med nize za opisovanje predmetnosti umeščamo tudi prislovna določila z deiktičnimi zaimki (***v tem primeru, na ta način***), ki zaradi svoje kohezivne vloge sicer predstavljajo sivo polje med nizi za opisovanje predmetnosti in organizacijo diskurza (prim. 2.3).

2.2. Nizi za vrednotenje

Nizi za vrednotenje oz. izražanje odnosa (angl. *stance expressions*) so nizi, s katerimi govorci razkrivajo svoj odnos do sporočanega in obenem vplivajo na naslovnikovo interpretacijo sporočil. Nekaj primerov:

- izražanje verjetnosti oz. resničnosti: ***naj bi se***
- izražanje mnenja: ***moram reči da, jaz mislim***
- izražanje negotovosti: ***ne vem, se mi zdi, mislim da, bolj ali manj***
- omiljevanje/ublažitev: ***na nek način, je v bistvu, bi rekel***
- izražanje modalnosti: ***lahko bi, moraš biti***
- izražanje dokaznosti: ***pravijo da, ni znano, po njegovem mnenju***
- poizvedovanje po mnenju: ***kaj mislite, se ti zdi***
- poudarjanje vsebine: ***je seveda, kaj šele, predvsem na, ali celo, še posebej***
- pozitivno ali negativno vrednotenje lastnosti: ***je fajn, zelo dobro, zelo pomembno***

Čeprav ima prednost interpretacija vloge niza kot celote, si pri določanju lahko pomagamo tudi s površinskimi kazalci, kot so prisotnost besedišča z modalnim pomenom (*treba, lahko, seveda, v bistvu, mogoče; moči, morati, smeti*) ali funkcijo (*naj, res*), glagoli mišljenja, hotenja ali zmožnosti (*vedeti, misliti, upati, želeti*). Pozor, to ne vključuje bi ali osebnih zaimkov.

Pri presojanju funkcije ustaljenih odzivov, ki so posredno vedno tudi v vlogi upravljanja govornje interakcije (2.3), se o njihovi funkciji odločamo na podlagi vsebine odziva, npr. funkcija vrednotenja v odzivih ***to je v redu, ja vem, se strinjam*** ali funkcija predmetnosti v vprašanih *kako ste, kje si*. Enako velja tudi za ustajene začetke stavkov, ki delujejo povezovalno, a obenem razkrivajo govornčev odnos do sporočanega, npr. ***zanimivo je da, pomembno je da*** (= vrednotenje).

2.3. Nizi za upravljanje diskurza

Nizi za upravljanje oz. organizacijo diskurza (angl. *discourse organizing expressions*) so nizi, s katerimi govorniki svoja sporočila oblikujejo v koherentno celoto in jih usklajujejo z drugimi okoliščinami sporazumevalnega dogodka. Nekaj primerov:

- (meta)besedilno povezovanje: ***se pravi, zaradi tega ker, glede na to da***
- tematska organizacija: ***kar se tiče, no v glavnem***
- povezovanje z naslovnikom: ***ja ja ja, a ne, veš kako je***
- zapolnjevanje vrzeli: ***kaj je že, kako se reče***
- vpljudnostne fraze: ***dobro jutro, dame in gospodje***

Čeprav ima prednost interpretacija vloge niza kot celote, si pri določanju lahko pomagamo tudi s površinskimi kazalci, kot so prisotnost besedišča z diskurzno funkcijo, kot so besedilni povezovalci (*sicer, namreč, tako da, v glavnem*). Pri pomensko bolj izpraznjenih diskurznofunkcijskih izrazih, kot so tipično govorniki pragmatični členki (*veš, a ne, ne, ja, no, pač, pa*), pri interpretaciji upoštevajmo tudi kontekst.

Kot besedilne povezovalce obravnavamo samo izraze, ki vzpostavljajo pomenska razmerja med deli besedila (npr. sosledica, protivnost, dodajanje, vzrok, pogoj; povzemanje, pojasnjevanje ...), ne pa drugih vrst medstavčnih veznikov (*s katerim, o katerih*).

Dvoumnost pri določanju pragmatične funkcije

Številni nizi v jeziku opravljajo več funkcij hkrati, zato njihovo določanje ni vedno preprosta naloga. V takih primerih nizu pripišemo funkcijo, za katero na podlagi analize primerov rabe ocenimo, da je najpogostejša oz. prevladujoča. To najlažje razberemo iz tipičnih kontekstov, v katerih se ta niz pojavlja, in parafrazo govornikov namer: »Govorec/pisec je ta niz besed uporabil, da bi ...«.

Še zlasti večfunkcijski so strukturno nezaključeni nizi, ki vsebujejo več ločenih skladijsko-pomenskih enot ali njihovih fragmentov, kot so kombinacije besed ali zvez z diskurzno vlogo in stavčnih fragmentov (npr. ***ja to je, ampak to je, tako da se***). Če se ni mogoče odločiti za eno samo funkcijo, damo prednost diskurzni ali vrednotenjski interpretaciji pred predmetnostno.

3. SLOVARSKA RELEVANTNOST

Z vidika slovarske relevantnosti nize razvrščamo na (i) **slovarsko relevantne** in (ii) **slovarsko nerelevantne** nize.

3.1. Slovarsko relevantni nizi

Kot slovarsko relevantne označujemo nize, ki predstavljajo ustaljene večbesedne enote z določeno pomensko ali slovnično vlogo, kakršne bi pričakovali tudi v različnih razdelkih splošnega razlagalnega slovarja, npr. kot samostojne iztočnice ali kot tipične besedne zveze znotraj posameznih iztočnic in njihovih pomenov. Običajno so to strukturno zaključeni nizi, ni pa nujno.

V ožjem smislu mednje uvrščamo vse ustaljene nize, katerih pomen oz. funkcija v jeziku nista povsem predvidljivi na podlagi pomenov posameznih besed. Tipični primeri so:

- frazeološke enote, tj. zveze z ekspresivnim ali metaforičnim pomenom: **tako rekoč, dame in gospodje, konec koncev**
- ustaljene odzivi ali vprašanja v govorni interakciji: **to je to, daj no, a veš, ja ne vem, ni problema, ne še, mislim da; kaj to pomeni, kako si, še kaj, zakaj pa**
- stalne zveze, tj. zveze s samostojnim pomenom brez ekspresivne konotacije, vključno s terminologijo: **javni sektor, nacionalni interes**
- zveze s skladenjsko oz. besedilno funkcijo, npr. **zaradi tega ker, kljub temu, tako da, še posebej če, po drugi strani, sicer pa, ali pa; in tako naprej, ali nekaj takega**
- ustaljeni začetki stavkov z besedilno funkcijo (**kar pomeni da,**) ali ustaljeni začetki vprašanj (**kaj pa če**)
- druge ustaljene skladenjske zveze, zlasti v vlogi prislovnih določil (**še zmeraj**), prilastkov (**bolj ali manj**) ali njihovih delov
- zveze povedkov in prostega morfema si/se: **zdi se, se spomniš, se dogaja**
- zveze predmetnopomenskih besed s tipičnimi predlogi (**priti do, iti za, govoriti o; pravica do, priložnost za; odvisno od, skupaj z; eden od, eno od**)
- oblikovno ustaljene predložne zveze s prislovno (**v povprečju, na prostem, do sedaj, v gosteh**), diskurzno (**poleg tega, s tem, pri tem**) ali naklonsko funkcijo (**z veseljem, na srečo, po mojem, v resnici, po svoje**)

Na spodnji meji slovarsko relevantnih zvez so kolokacije, tj. pogoste skladenjsko zaključene sopojavitve besed, ki zaradi svoje pomenske transparentnosti ne zahtevajo posebnega pomenskega opisa, a v slovarju funkcionirajo kot tipične ponazoritve pomenov besed, ki jih sestavljajo (npr. **prehodno stanje, v letošnji sezoni**).

Kot slovarsko relevantne obravnavamo tudi kolokacije s pomensko izpraznjenimi kolokatorji oz. zaimki (npr. **vsako leto, nekaj let, vsaki dan; zelo pomembno, zelo težko; tule gor, te zadeve, vse ostalo, vsi skupaj**) in števnikih (**dvajset let, dva meseca, tri leta, tri dni, dva tisoč, pol sedmih, ob petih**) ter strukturno zaključene predložne zveze (**iz tujine, v soboto, po telefonu, na fotografiji, z glavo, v finalu; v času, pred začetkom, po poročanju, z namenom, na osnovi**), z izjemo zaimkovnih predložnih zvez, izpostavljenih v razdelku 3.2.¹ Kot kolokacije obravnavamo tudi dele predložnih zvez brez predloga (**svetovnem prvenstvu, današnji dan, mojem mnenju**).

¹ S tem se zamejitev kolokacij v tej nalogi razlikuje od semantično usmerjenih opredelitev, denimo znotraj projekta KOLOS, kjer kot kolokacije izločajo zveze s pomensko manj relevantnimi modifikatorji. To širšo opredelitev v tej fazi dopuščamo, ker (1) že v izhodišču operiramo s statistično najbolj izstopajočimi zvezami v

3.2. Slovarsko nerelevantni nizi

Kot slovarsko nerelevantne označujemo vse ostale nize, za katere menimo, da nimajo nobene ustaljene pomenske ali slovnične vloge v jeziku in jih zato ne bi pričakovali v slovarskih priročnikih. Gre torej za proste besedne zveze ali strukturno in pomensko nezaključene nize med katerimi prevladujejo strukturno nezaključeni fragmenti:

- stavkov (**da gre za, predlagam da, je bila v, pri čemer je, ne bo več; ali pa če**),
- daljših besednih zvez (**devet pet osem, koli že**),
- predložnih zvez: **s tistim, v zadnjih, po drugi**
- razstavljivi nizi funkcijskih besed: **ja tako da**

Med strukturno zaključenimi zvezami kot nerelevantne označujemo tudi naslednje skupine:

- lastna imena (**Evropska unija, Radio Center, Tina Maze**) in njihovi deli (**prvi interaktivni multimedijski portal**)
- zloženi povedki za izražanje glagolskih časov ali modalnosti: **bo šel, bi lahko, je šel, morajo biti, ne smemo, sem gledal**. (Pozor, če je zloženi povedek razvil svoj samostojni pomen oz. funkcijo, ga označujemo kot slovarsko relevantnega, npr. ustajene fraze *boš videl, bi rekel, moram povedati, moram iti, je rekel/sem rekla*. Enako velja za slovarsko relevantne zveze povedka in prostega morfema *si/se*, npr. *se dogaja, se spomniš*).
- predložne zveze z deiktičnimi zaimki: **pri nas, z njimi, o njem, k temu, o katerem, s kom ...**
- dele ustaljenih predložnih besednih zvez (**skladu z**), ki se običajno ne pojavljajo v drugih vlogah ali zvezah

Pri nekaterih skupinah izrazov se o slovarski relevantnosti odločamo po presoji glede na ustaljenost v rabi ter skladenjsko in pomensko razstavljivost:

- ustaljeni odzivi in vprašanja: kot nerelevantne označujemo pomensko in skladenjsko razstavljive izjave (**ne morem, jaz ne bom; kje si, kaj pa je to**). Za primere slovarsko relevantnih vprašanj in odzivov (npr. *daj no, ja ne vem, zakaj pa*) glej 3.1.
- ustaljeni začetki stavkov: kot nerelevantne označujemo stavčne fragmente brez posebne slovnične funkcije (**predlagam da, verjamem da, dejstvo je da, zdi se da**). Za primere slovarsko relevantnih stavčnih fraz z diskurzno-povezovalno funkcijo glej 3.1 (npr. *kar pomeni da, kaj pa če*).

jeziku, (2) tudi pomensko izpraznjeni kolokatorji razkrivajo nekatere pomembne semantične lastnosti jeder, ob katerih se pojavljajo, in (3) so take zveze kot slovarsko relevantne brez specifičnih navodil v prvem koraku prepoznavali tudi označevalci.

Dvoumnost pri določanju slovarske relevantnosti

Presoja slovarske relevantnosti je v določeni meri subjektivna, pomembno pa je, da pri odločanju razmišljamo o najširšem krogu slovarskih uporabnikov. Med njimi so denimo tudi učenci slovenščine kot tujega jezika, ki v slovarjih pogosto iščejo tudi izraze, ki se rojenim govorcem zaradi svoje pogostosti ali pomenske izpraznjenosti zdijo povsem samoumevni (npr. [in tako naprej](#), [no ja](#), [ja ja](#), [mislim da](#), [kaj pa](#) itd.). V primeru dvoma niz raje označimo kot relevanten.

POZOR! Za razliko od opisovanja zgradbe in funkcije, kjer se opiramo na kriterij najpogostejše oz. prevladujoče rabe, pri presoji slovarske relevantnosti niz označimo kot slovarsko relevanten ne glede na to, kako pogosta je raba niza v drugih vlogah.

Taka sta denimo niza *to je* in *še nekaj*, ki se večinoma pojavljata kot (slovarsko nerelevantna) strukturno nezaključena fragmenta, a tudi kot slovarsko relevantna besedilna povezovalca, zato ju označimo kot slovarsko relevantna:

[nerelevantno] <i>To je presenečenje.</i>	[relevantno] <i>Uresničuje svoje sanje, to je postati plesalka.</i>
[nerelevantno] <i>Kupi še nekaj pomaranč.</i>	[relevantno] <i>Še nekaj: prideš v petek?</i>

POTEK DELA

Vsak označevalec prejme svojo Excelovo datoteko s tabelo, v kateri je pet stolpcev:

1. Besedni niz
2. Polje za opis skladišne strukture
3. Polje za opis pragmatične funkcije
4. Polje za vpis slovarske relevantnosti
5. Povezava do naključnega vzorca primerov rabe v korpusu
6. Polje za morebitne komentarje

Ko zaključite z analizo dodeljenih primerov (tj. zapolnitvijo drugega, tretjega in četrtega stolpca), oddano datoteko pošljete na dogovorjeni elektronski naslov s pripisom porabljenega časa.