

Slovar tviterščine

Polona Gantar, Iza Škrjanec, Darja Fišer, Tomaž Erjavec

Konferenca JTDH
Filozofska fakulteta UL, 30. 9. 2016

Pregled

1. Namen
2. Korpus tвитov
3. Luščenje iz korpusa
4. Kategorizacija in izbor leksike
5. Zgradba slovarskega gesla
6. Slovarski vmesnik Lexonomy
7. Prihodnje delo

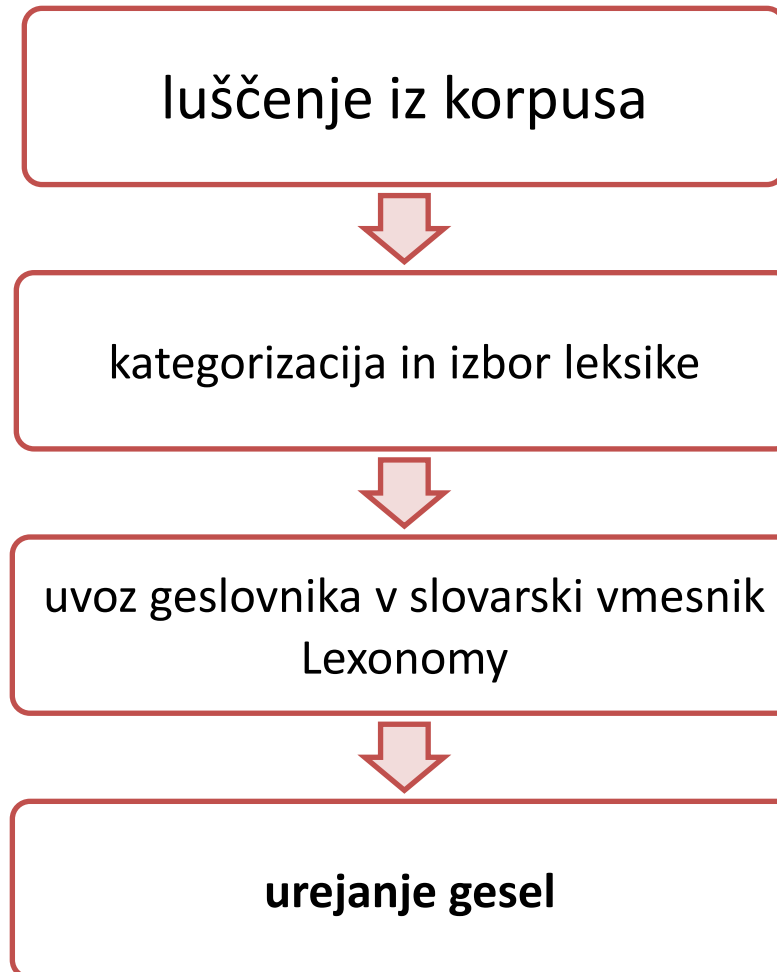
Namen

- zbirke besedišča iz računalniško posredovane komunikacije
 - *The Internet Dictionary* (Crumlish et al., 1995)
 - *A glossary of netspeak and textspeak* (Crystal, 2004)
- spletna slovenščina
 - raziskave posameznih fenomenov jezika
 - **celovit in sistematičen opis besedišča**
- namen:
 - seznaniti uporabnike z najbolj **značilnim besediščem** slovenskih tвитov (primeri iz žive rabe)
 - leksikalne analize jezikovnega **(ne)standarda** in načina vključevanja **tujih besed** v slovenščino

Korpus tvitov

- del korpusa **Janes** (uporabniške spletne vsebine)
- korpus **Tviti v0.3.4** (Erjavec et al., 2015)
 - velikost: več kot 56 milijonov besed (4 milijone tvitov)
 - čas objave tvitov: 2011–2014
 - metapodatki: o avtorju (**zasebni in korporativni računi**) in o besedilu
 - avtomatska mera **jezikovne standardnosti**: od 1 (standardno) do 3 (zelo nestandardno) (Ljubešič et al., 2015)
 - raba slengizmov in narečnih besed, besedni red, fonetizirani zapis, slovnična pravila

Postopek izdelave slovarja



Luščenje iz korpusa – izdelava geslovnika

- tviti v jezikovno **nestandardni slovenščini** (L2 in L3) uporabnikov **zasebnih računov** → podkorpus
- frekvenčna seznama lem in pojavnic: primerjava
 - s **celotnim korpusom Janes**
 - s **korpusom Gigafida**
- zastopanost v **SSKJ 1** in **SNB**

Kategorizacija in izbor leksike

nestandardna beseda • nova beseda (tuje nepodomačeno, tuje podomačeno) • krajšava • nov pomen

- ožji izbor leksike
- informacija za uporabnike

Nestandardnost: zapis in beseda

nestandardna beseda • nova beseda (tuje nepodomačeno, tuje podomačeno) • krajšava • nov pomen

- besede in zapise besed, ki jih ni mogoče pričakovati v besedilih, ki predstavljajo gradivno osnovo za **standardizacijski opis jezika**
- nestandardnost **zapisa**: fonetiziran zapis sicer standardnih besed: *jaz* (*jz, js, jest, jas*) → zajeto v širši geslovník, vendar ne v ožji izbor

Nestandardnost: zapis in beseda

nestandardna beseda • nova beseda (tuje nepodomačeno, tuje podomačeno) •
krajšava • nov pomen

- nestandardnost **besede**:
 - besede s **standardno ustreznico**: *izležavati (se) – lenariti, poležavati; švoh – šibek*
 - besede **brez standardne ustreznice**: *štopati, odfuk*
 - besedotvorne posebnosti:
 - **izbira obrazila**: *fotoaparat – fotkič, fotič*
 - zaznamovanost v SSKJ 1 in SNB

Nova beseda

nestandardna beseda • ***nova beseda (tuje nepodomačeno, tuje podomačeno)*** •
krajšava • *nov pomen*

- nova beseda:
 - **slovenska** osnova: odslediti
 - iz **tujega jezika**: stopnja vključenosti v slovenščino
 - **podomačeni zapis**: *thanks – tenks, comp(uter) – komp*
 - **pregibanje**: *follow – followati* („Me lahko followaš, da ti pošljem DM?“)
 - tuje **nepodomačeno** (okolje!): „thanks ja, sem vidla“

Krajšava

*nestandardna beseda • nova beseda (tuje nepodomačeno, tuje podomačeno) •
krajšava • nov pomen*

- kratice, krajšave in alfanumerični nizi
 - *ZS, np, btw, app, omg/omfg/omb, gr8*

Nov pomen

*nestandardna beseda • nova beseda (tuje nepodomačeno, tuje podomačeno) • krajšava • **nov pomen***

- nov pomen – **pomenski premik** glede na obstoječe opise v SSKJ 1 in SNB
 - štekati (se): *pogovorno razumeti (se) v SNB*
 - štekati: občasno prenehati delovati za krajši čas, predvsem v zvezi z elektronskimi napravami in programsko opremo („a še komu šteka Firefox za popizdit“)
 - *pičiti, koma, sledilec*

Zgradba gesla in slovarski vmesnik Lexonomy

- primeri potrjenih gesel: *afkors, komad, thx, izležavati, freelancer*
- slovarski vmesnik **Lexonomy** (Měchura, 2012):
 - prosto dostopen
 - uvoz in izvoz slovarske baze
 - fleksibilnost pri nastavitvi zgradbe gesla
 - testni slovar: <http://lexonomy.cjvt.si/>

V prihodnje

- 1021 gesel:
 - 22 gesel potrjenih, ostala v fazi urejanja
- izbor dobrih primerov: aplikacija **GDEX** (Kosem et al., 2013)
- **povezave:**
 - med variantami v slovarju (*kafe* → *kofo, kofi, kufet*)
 - na geslo v SSKJ 1 in SNB
- **integracija v druge slovarske baze** (za izdelavo Slovarja sodobne slovenščine, Gorjanc et al., 2015)

thanks tenks tenkju ty tnx thx huala hwala hvala

