
OZNAČEVANJE UDELEŽENSKIH VLOG V UČNEM KORPUSU ZA SLOVENŠČINO

SIMON KREK, POLONA GANTAR, KAJA DOBROVOLJC IN IZA ŠKRJANEC

Jezikovne tehnologije in digitalna humanistika, Ljubljana, 29. 9 – 1. 10. 2016

ZGODOVINA

- XLike - Cross-lingual Knowledge Extraction; IJS (2012-2014)
- Clarin Slovenija - Slovenska raziskovalna infrastruktura za jezikovne vire in tehnologije (2015)
- CJVT – Center za jezikovne vire in tehnologije (2016)
- Bilateralni projekt med Slovenijo in Hrvaško (2015-2017)

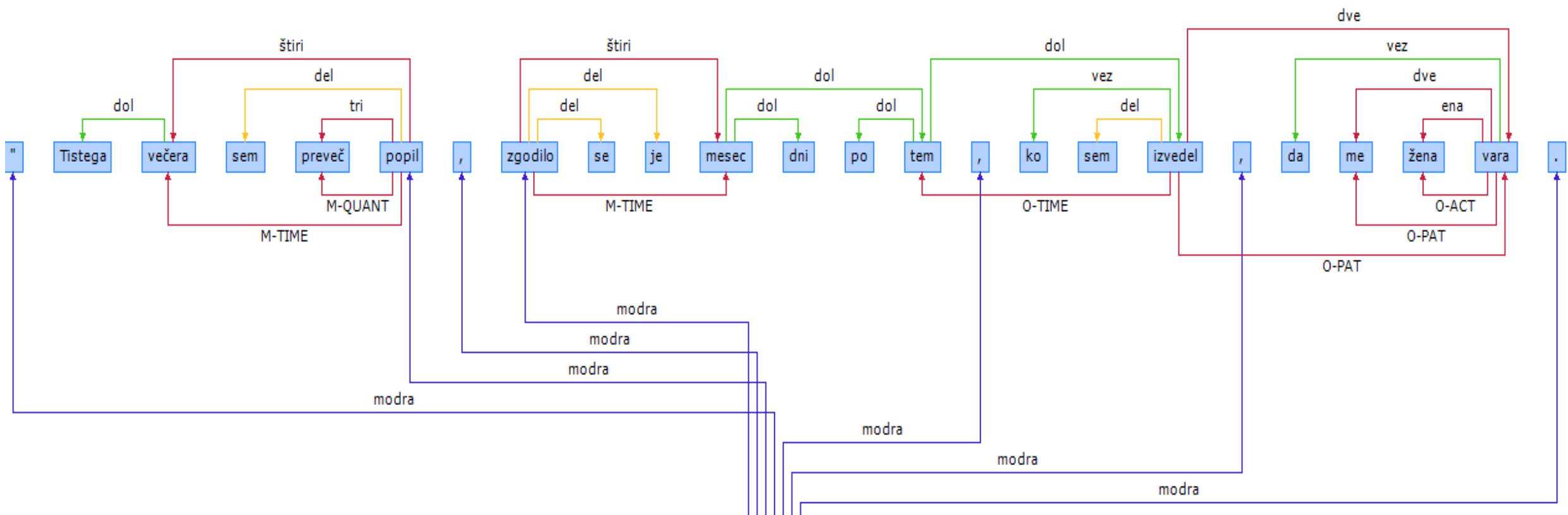
- Sodelujoči : Simon Krek, Kaja Dobrovoljc, Sara Može, Polona Gantar, Iza Škrjanec Dafne Marko in ...
- Dostopnost: Clarin.si; CC

CILJI IN METODE

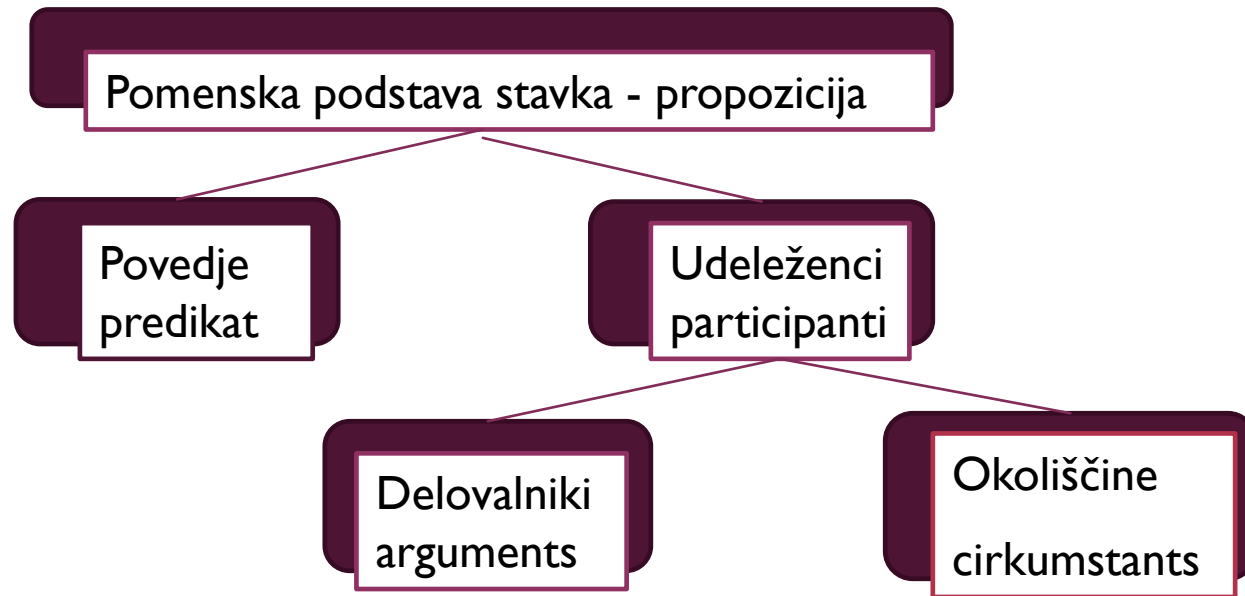
- Izhajati iz jezikovnih specifik slovenščine in upoštevati medjezikovno kompatibilnost oznak in mednarodne standarde (→ UD)
- Zagotoviti optimalen nabor kategorij
- Upoštevati odsotnost leksikona glagolske vezljivosti
- Metode semantičnega označevanja in semantične kategorije
 - Prague Dependency Treebank (Mikulová et al., 2006)
 - PropBank (Palmer et al., 2005)
 - Verbnet (Kipper et al., 2006)
 - AnCora (Taulé et al., 2011)
 - SoNaR (Schuurman et al., 2010)
 - nabor oznak za hrvaščino (Filko et al., 2012)
 - češki valenčni leksikon Vallex in FrameNet (Backer et al., 1998)

KORPUS IN OZNAČEVALNIK

- SSJ500k
- SentenceMarkup System (Dobrovoljc et al., 2012)



TEORETIČNA IZHODIŠČA



Udeležence in njihove udeleženske vloge določa prototipična prepozicija za dani pomen glagola, ki se uresničuje v njegovi vezljivostni shemi.

pomensko obvezni

pomensko neobvezni – PropBank, Verbnet

pomensko obvezni

pomensko neobvezni - PDT

NABOR UDELEŽENSKIH VLOG - DELOVALNIKI

Oznaka	Opis		Primer
ACT	aktant, vršilec	delujoči udeleženci, povzročitelji ali nosilci dejanja	<u>Oče</u> ACT dela.
PAT	prizadeto	prizadeti predmet dejanja	Kuhajo <u>kosilo</u> PAT
REC	prejemnik	prejemnik, posredni udeleženec dejanja; nedelovalniški udeleženec, ki mu je dejanje v škodo ali v prid	<u>Prijatelju</u> REC sem poslal darilo.
ORIG	izvor	izhodišče, izvor/vir/povod dejanja	Plašč je dobil od <u>očeta</u> ORIG
RESLT	učinek	učinek, rezultat, cilj dejanja	Imenovali so ga za <u>predsednika</u> RESLT

NABOR UDELEŽENSKIH VLOG – OKOLIŠČINE: ČAS IN PROSTOR

Oznaka		Opis		Primer
ČAS	TIME	čas	konkretni trenutek ali interval dejanja; kdaj	Med <u>počitnicami</u> _{TIME} ni niti enkrat posijalo sonce.
	DUR	trajanje	trajanje stanja, dejanja koliko časa	Prišel je za en <u>mesec</u> _{DUR}
	FREQ	pogostnost	frekvenca dejanja; kako pogosto, kolikokrat	Vsak <u>dan</u> _{FREQ} se mučimo s tem.
PROSTOR	LOC	kraj	konkretna lokacija, kraj, mesto dejanja; kje	V <u>bližini vasi</u> _{LOC} stoji kozolec.
	SOURCE	začetna lokacija	začetna točka v prostoru; od kod	S <u>stropa</u> _{SOURCE} odpada omet.
	GOAL	končna lokacija	končna točka v prostoru; kam	Prišel je <u>domov</u> _{GOAL}

Oznaka		Opis		Primer
VZROČNOST	AIM	namen	namen dejanja; čemu, s kakšnim namenom	Telovadi, da bi <u>shujšala</u> _{AIM}
	CAUSE	vzrok	vzrok dejanja; zakaj	Umrl je zaradi <u>srčne kapi</u> _{CAUSE}
	CONTR	protivnost	nepričakovana posledičnost dejanja; kljub čemu	<u>Medtem ko se plače nižajo</u> , _{CONTR} cene rastejo
	COND	pogojnost	pogoj za obstoj dejanja ali dogodka	<u>Če neha deževati</u> _{COND} gremo na kavo.
	REG	ozir	glede na, primerjava	Preiskava je potekala <u>v skladu z zakonom</u> _{no-REG}
NAČIN	ACMP	spremljavo	predmet, oseba ali dogodek, ki spremlja dejanje ali druge udeležence	Mama je s <u>sinom</u> _{ACMP} odšla v cirkus.
	RESTR	omejitev	izjema, omejitev	Vsi so bili tam razen <u>tebe</u> _{RESTR}
	MANN	način	načinovna lastnost dejanja,	Dela <u>prepočasi</u> _{MANN}
	MEANS	sredstvo	sredstvo ali orodje za izvedbo dejanja	Piše s <u>peresom</u> _{MEANS}
KOLIČINA	QUANT	količina	količina, razlika	Cena goriva se je podražila za 3 <u>cente</u> _{QUANT}

NABOR UDELEŽENSKIH VLOG – GLAGOLSKE ZVEZE

Oznaka	Opis	Primer	
GLAGOLSKE ZVEZE	MWPRED	zveze z nedoločniki	Dati vedeti _{MWPRED}
	MODAL	zveze biti + modalnega prislova/pridevnika	je treba, je mogoče
	PHRAS	pomensko neprozorne zveze	liti na živce _{PHRAS}

biti + pridevnik = **RESLT**: stvar je zapletena_{RESLT}

Imeti avto – PAT

biti + samostalnik = **PAT**: območje medenice je središče telesa_{PAT}

Imeti krompir/smolo – PHRAS

biti + prislov = **MANN**: da vam bo toplo_{MANN}

Imeti prednost/možnost/predavanje - LVC

KONKURENČNA POMENSKOSKLADENJSKA RAZMERJA - I

■ Med delovalniki

- ACT – PAT; ACT – REC; PAT – REC; PAT – RESULT; PAT – ORIG

■ Med delovalniki in okoliščinami

- ACT – LOC ...
- PAT – MEANS ...
- REC – AIM ...

■ Med okoliščinami

- TIME – LOC; TIME – MANN ...
- LOC – RESULT; LOC – MANN ...
- CAUSE – LOC; CAUSE – TIME; CAUSE – MANN

Pripis udeleženske vloge mora ostati nespremenjen v različnih skladenjskih realizacijah udeleženca.

AKTANT – PRIZADETO

AKTIV

dogodki (ACT) so se odvijali

* ~~Kdo~~ (ACT) odvijaja dogodke (PAT)

Kdo (ACT) malce zaplete stvar (PAT)

PASIV

Pozitivna diskriminacija (PAT) se označuje kot privilegij

Kdo (ACT) označuje pozitivno diskriminacijo (PAT) kot privilegij

Stvar (PAT) je malce zapletena.

Stvar (PAT) se je malce zapletla

PROSTORSKO IZRAŽENI DELOVALNIKI

- ACT – LOC
 - v bolnišnici (LOC → ACT) bodo uvedli
 - pri Fujifilmu (LOC → ACT) so objavili
 - na centru (LOC → ACT) se bojo lotili
 - so sporočili iz združenja (LOC → ACT)
 - v Avstrijskem časopisju (LOC → ACT) objavljajo
 - v osnovni šoli v Bistrici (LOC → ACT) pripravljajo
- REC – LOC
 - je zganila v meni (LOC → REC)
 - boste v sebi (LOC → REC) začutile življenje
 - bodo povzročile nevšečnosti na televiziji (LOC → REC)

ČASOVNOPROSTORSKA IN VZROČNA POMENSKOSKLADENJSKA RAZMERJA

- TIME – LOC = **EVENT**
 - sploh ni padel v vojni (TIME – LOC → EVENT)
 - na enem od sestankov (TIME – LOC → EVENT) jim je dejala
- TIME – CAUSE
 - dekle je ob vzvratni vožnji (TIME → CAUSE) začelo vpiti
 - ob zadnjih zapletih (TIME → CAUSE) se je županja odločila

V PRIHODNJE

- Na podlagi nabora udeleženskih vlog in meril označiti še preostali del polovice skladijsko označenega korpusa SSJ500k (4,500 stavkov)
- Prehod v Universal Dependencies
- Obligatornost udeleženskih vlog

- Dodati nov nivo za označevanje večbesednih enot (MWE)

Zahvaljujem se vam_[REC] za pozornost_[PAT]!