

Quotations, Coreference Resolution, and Sentiment Annotations in Croatian News Articles: An Exploratory Study

Jelena Sarajlić, Gaurish
Thakkar, Diego Fernando
Válio Antunes Alves and
Nives Mikelić Preradović

jelenasarajlic2@gmail.com
gthakkar@m.ffzg.hr
dfvalio@ffzg.hr
nmikelic@ffzg.hr

Faculty of Humanities and Social Sciences
University of Zagreb
Zagreb 10000
Croatia

Overview

1

Introduction; Quotes & Quoting

What was our motivation?

What are quoting types or styles?

2

Annotation Process

How did our annotation process go?

What does the tool look like?

3

Description of Problems

What are the problems we ran into and why are they important?

Introduction

- Extraction of quotes can greatly contribute to **stance analyses, tracking** the general public's or a specific person's **change of opinion in time**, and can be useful to numerous issues in **digital humanities**.
- Croatian = **low resource language**
- We chose a small set of 140 documents, **120 of which contained direct quotes**, of the **Croatian** part of the **SETimes corpus** and **annotated them in terms of direct quotations, their parts and sentiments**.
- SETimes is a parallel corpus based on the contents published on the SETimes.com news portal and covers **10 languages: English, Croatian, Bulgarian, Bosnian, Greek, Macedonian, Romanian, Albanian, Serbian and Turkish**.

Types of quotes

direct

Ana said: "I
love cookies!"

indirect

Ana said
that she
loves
cookies.

mixed

Ana said that
she "loves
cookies".

Different types of (direct) quotes

"I'm really happy for Fabio" Materazzi **told** the Apcom news agency Friday. *"I feel part of this distinction because I think that all the Azzurri helped a great champion like Cannavaro win an important prize"*

1

Michelle Mcgrath **said**, *"We stand ready to support you in every way"*.

2

"We stand ready to support you in every way" Blair **said**.

3

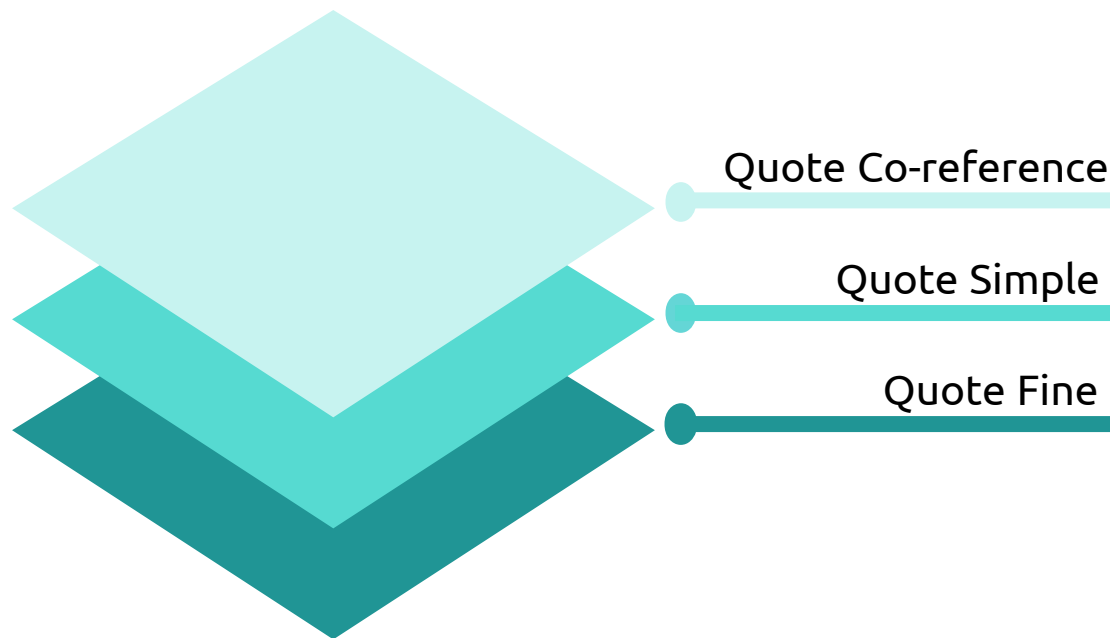
Tony Mcgrath visited Iraq... *"We stand ready to support you in every way"* the Prime Minister **said**.

4

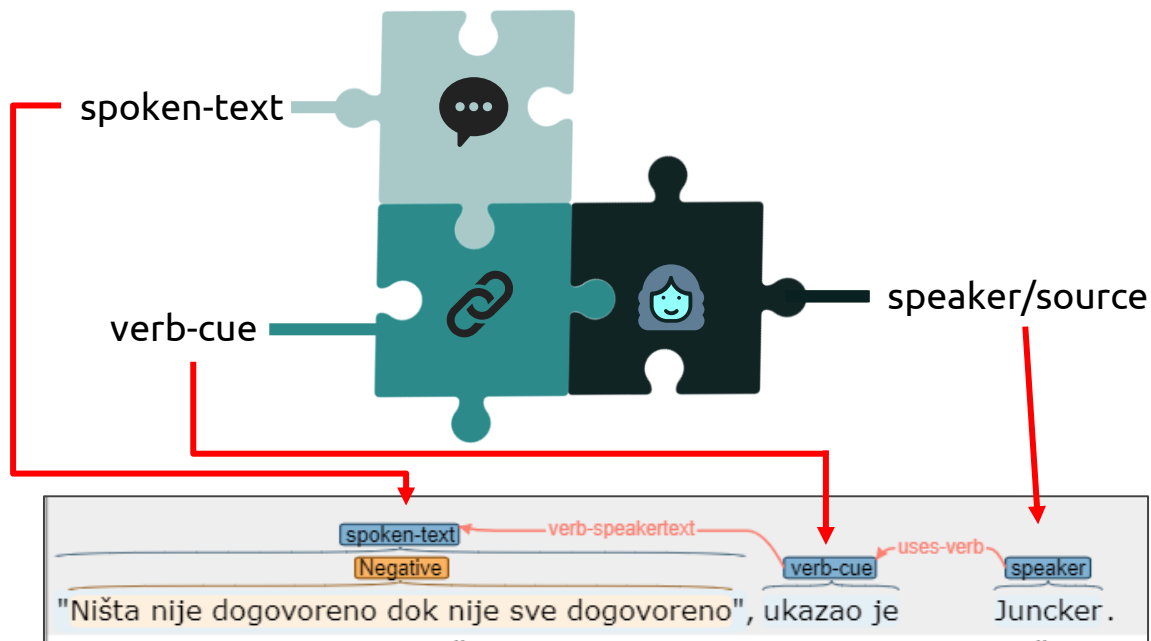
Tony Mcgrath visited Iraq... He **said**, *"We stand ready to support you in every way"*.

5

Annotation – three custom layers

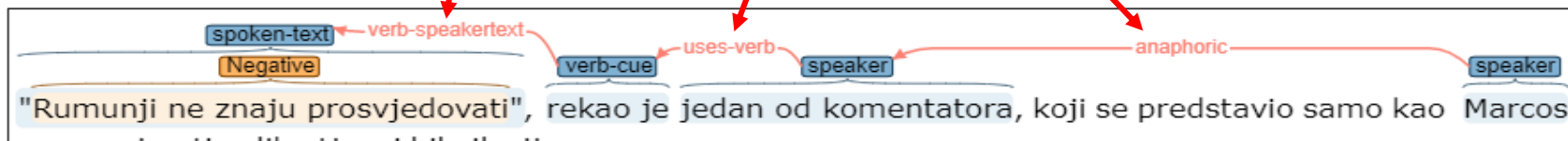
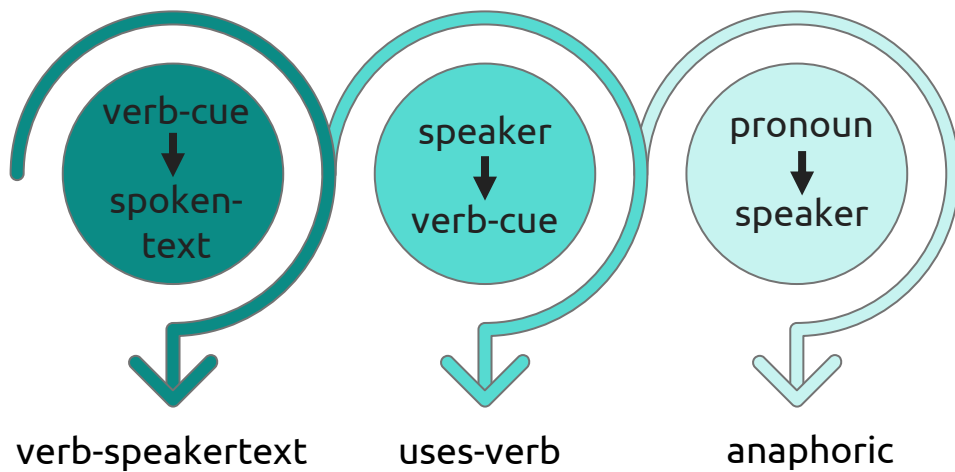


Quote Fine layer – quote parts



"Nothing is agreed until everything is agreed," Juncker stressed.

Quote Co-reference layer – relations between quote parts



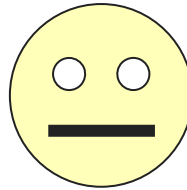
"Romanians do not know how to protest," said a commenter identified only as Marcos

Quote Simple layer – the sentiment

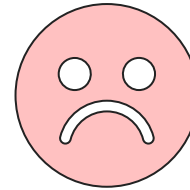
sentence (or clause) level



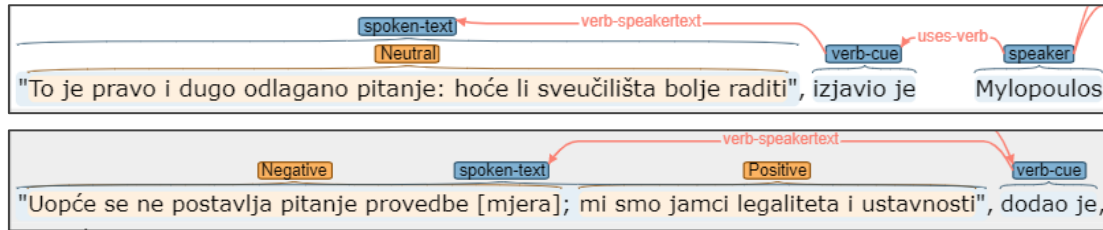
positive



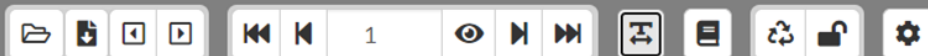
neutral



negative



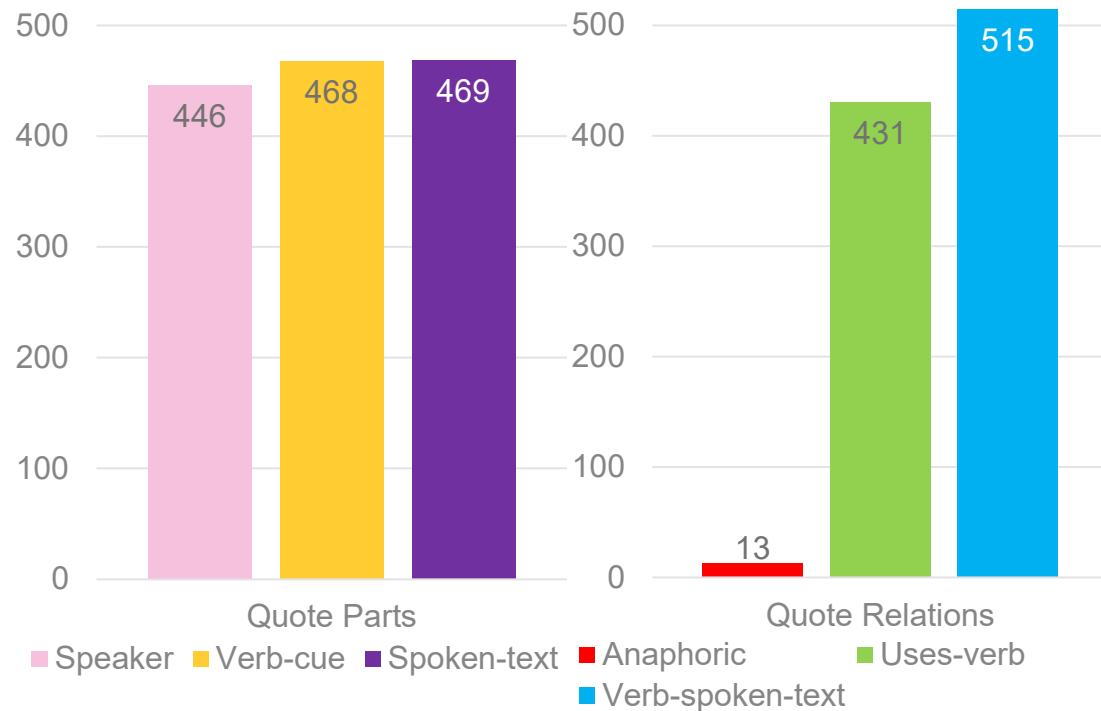
"That's the real and long-delayed question: will the universities do a better job", Mylopoulos stated...
"It is not at all a question of the implementation [of the measures]; we are merely the guarantors of legality and constitutionality", he added...



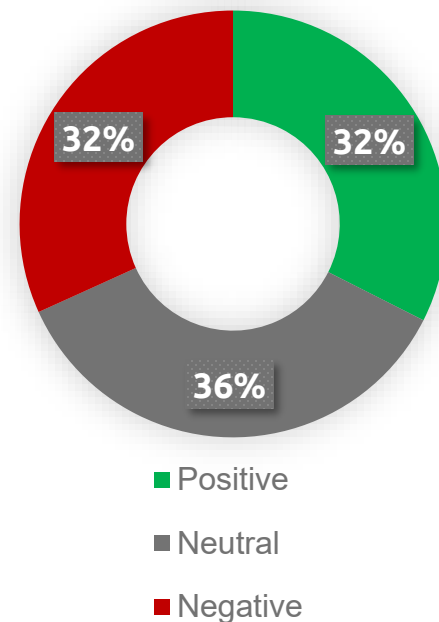
9. Na temelju ranijeg zakona postalo je čisto kako je za prodaju nekretnosti prema staroj sudskoj proceduri trebalo izmisliti do pet godina, dok će sa
10. Prema Parvodićevim riječima, mjera će potaknuti sigurnost kreditora i dati im puno više slobode u odobravanju kredita.
11. "Drugim riječima, znatno više ljudi bit će u mogućnosti dobiti kredite, što će dovesti do manjih troškova kredita".
12. Srbija, čiji su napori u provedbi gospodarskih reformi pohvaljeni u izvješću Svjetske banke ranije ove godine, ističe kako su provedba novih regulativa o prodaji kompanija -- i otpisivanje njihovih dugova državi -- rezultirali ubrzanjem privatizacijskih sporazuma u posljednjih šest mjeseci.
13. Prema riječima ravnatelja Agencije za privatizaciju Miodraga Đorđevića, dobit Srbije od privatizacije 501 kompanije ove godine iznositi će 390 milijuna eura.
14. Dogovorena ulaganja donijet će još 118,2 milijuna eura.
15. Od ukupnog broja privatiziranih kompanija ove godine, 204 su prodane putem aukcije, 15 na natječaju, a 282 paketa državnih dionica prodano je na tržištu kapitala.
16. "U 2006 godini planiramo prodaju 300 kompanija na dražbi, raspisivanje 30 natječaja i prodaju 250 paketa dionica iz lisnice Akcijskog fonda",
17. Među velikim kompanijama predviđenim za privatizaciju na natječaju nalaze se rafinerija nafte u blizini Beograda, velika pivovara i nekoliko špediterskih i transportnih kompanija.

Some statistics

Number of annotations in Quote Layers



Distribution of sentiment tags



Problems discovered while annotating

Sentiment tagging problem

Sentence parts differently annotated

Specific Croatian-related problems

Passive

Cross-branching

Quoting problems

Mixed quotes

Seems to be quoted but no quotes

Quotes marking something other than speech

Speaker/source problems

Speaker mentioned multiple times differently

Non-person speaker and Anonymous speaker

Apposition(s) alongside speaker

Quotes from parts of documents or a collective

“Position in text” problems

Quotes without an apparent speaker or verb-cue

Indirectly correlated speaker / verb-cue

Speaker and/or verb-cue in the middle of the quoted text

Indirectly correlated speaker / Indirectly correlated verb-cue

6. Croatian:

Sličnu
similar-Agpf_{say}

izjavu
statement-Nc_{fsa}

dala
give-Vmp-_{sf}

je
be-Var_{3s}

Sarah Lum,
speaker

predstavnicu
representative-Nc_{fsn}

Američkog
American-Agpm_{sgy}

ureda
office-Ncm_{sg}

u
in-Sl

Prištini.
Pristina-Np_{fsl}

“QUOTE”,
quote

kazala
say-Vmp-_{sf}
verb-cue

je.
be-Var_{3s}

tokens separating
speaker from her quote
and the corresponding
verb-cue

English: Similar comments came from Sarah Lum, representing the US office
in Pristina. “QUOTE,” she said.

Speaker and/or verb-cue in the middle of the quoted text

8. Croatian:

“QUOTE1.1”,
part 1/2 of the quote

rekao je
say-Vmp-sm be-Var3s
v e r b - c u e

krajem
end-Sg

rujna
September-Ncmmsg

za
for-Sa

SETimes
SETimes-Npmsan

Angel Kishev
speaker

80-godišnji
80-year- old-Agpmсны

ribar
fisherman-Ncmsn

iz
from-Sg

grada
town-Ncmsg

Chernomoretsa.
Chernomorets-Npmsg

“QUOTE1.2”
part 2/2 of the quote

English: “QUOTE1.1,” Angel Kishev, an 80-year-old fisherman from the town of Chernomorets, told SETimes in late September. “QUOTE2.2”

Quotes without an apparent speaker or verb-cue

9. Croatian:

Plan, plan-Ncmsn	dodao add-Vmp-sm	je be-Var3s	Holkeri, speaker	za for-Sa
cilj goal-Ncmsan	ima have-Vmr3s	osigurati ensure-Vmn	da that-Cs	ciljevi goal-Ncmpn
standarda standard-Ncmpg	koje which-Pi-mpa	su be-Var3p	odredili set-Vmp-pm	UN – UN-Npmpn
“učiniti make-Vmn	Kosovo Kosovo-Npnsa	boljom good-Agcfsiy	sredinom place-Ncfsi	za for-Sa
sve: everyone-Agpnsay	sigurnom, safe-Agpnsly	stabilnom stable-Agpmsly	i and-Cc	prosperitetnom” – prosperous-Agpmsly
postanu become-Vmr3p	realnost. reality-Ncfsa	“QUOTE” quote		

is this Holkeri's
quote?

English: The plan, Holkeri said, aims to ensure that the goal of the UN-endorsed standards – “to make Kosovo a better place for everyone: safe, stable and prosperous” – will become a reality. “QUOTE”

Quotes marking something other than speech

11. Croatian:

Novi new-Agpmsny	bi be-Vaa3s	se self-Px-sa	odbor committee-Ncmsn
izravno directly-Rgp	usredotočio focus-Vmp-sm	na on-Sa	žrtve victim- Ncfpa
dokumentirajući document-Rr	“približnu istinu”. approximate-Agpfsay truth-Ncfpa quotes marking something other than speech		quote?

English: A new panel would focus squarely on the victims in documenting an “approximation of the truth”.

Quotes from parts of documents or a collective

spoken-text
Positive
verb-speaker-text
verb-cue
"Vlada je i dalje stabilna i nastavlja pokazivati visok stupanj konsenzusa oko pridruživanja Europskoj uniji kao strateškom prioritetu", navodi se u izvješću.

"The government has remained stable and continues to express a high degree of consensus on accession to the European Union as the strategic priority," the report says.

is this a
valid source?

Non-person speaker / Anonymous speaker

15. Croatian:

“QUOTE”,
quote

kazalo je
say-Vap-sn be-Var3s

dvoje čelnika
two-Mls leader-Ncmpg

u priopćenju nakon sastanka.
in-Sl statement-NcnsL after-Sg meeting-NcmSg

English: “QUOTE” the two leaders said in a statement after the meeting.

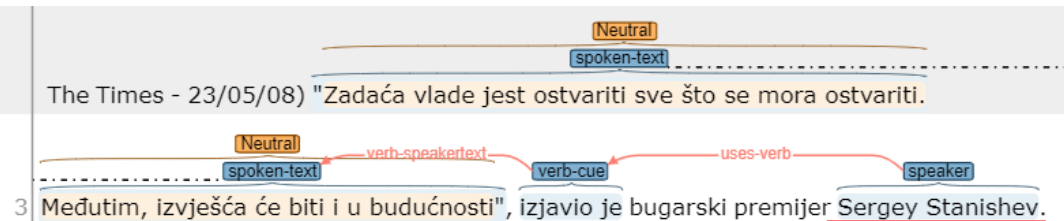
Seems to be quoted but no quotes

12. Croatian:

Čelnici	moraju	voditi,	a	ne
leader-Ncmpn	must- Vmr3p	lead-Vmn	and-Cc	not-Qz
samo	pratiti	svoje	pristaše”.	
only-Rgp	follow-Vmn	one’s own-Px-mpa	follower-Ncmpn	

English: Leaders must lead, and not merely follow their followers”.

Speaker mentioned multiple times differently

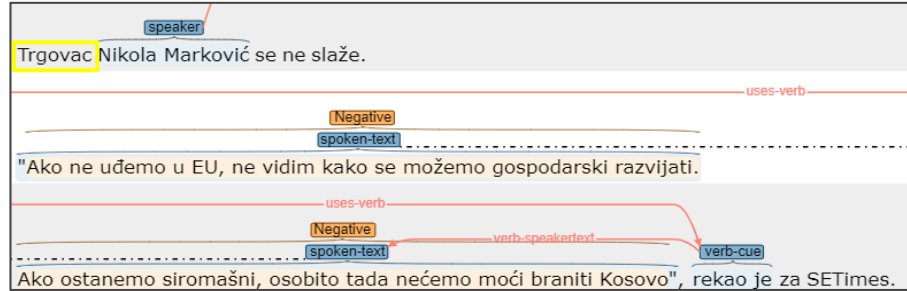


"The cabinet's task is to achieve everything that must be achieved. However, there will be future reports," Bulgarian Prime Minister Sergi Stanishev said.

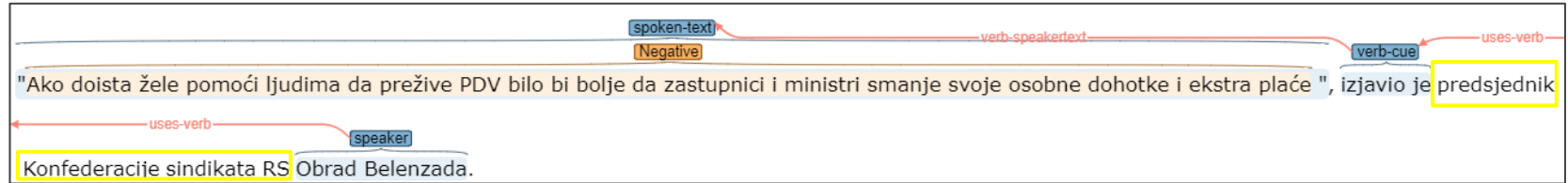


"It takes time to impose these new European standards," Stanishev said.

Appositions alongside speaker

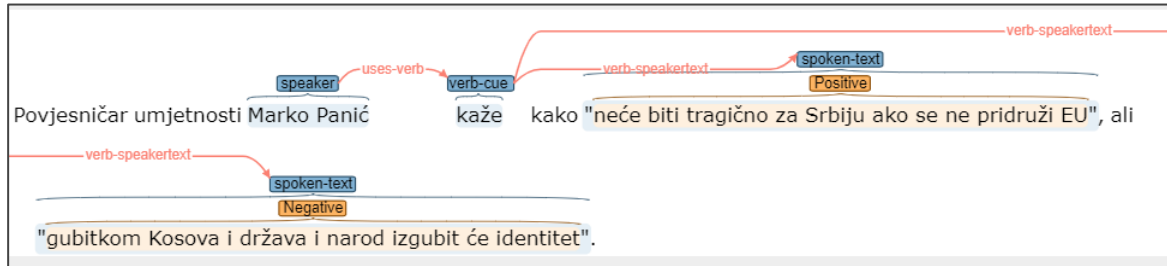


Salesman Nikola Markovic disagrees. "If we do not join the EU, I do not see how can we develop economically. If we remain poor, especially then we will not be able to defend Kosovo," he told SETimes.



"If they really want to help the people to survive VAT, it would be better for MPs and ministers to reduce their paychecks and extra salaries," says **RS Union Confederation President** Obrad Belenzada.

Mixed quotes



Art historian Marko Panic says that "it will not be tragic for Serbia not to join the EU," but "by losing Kosovo both the state and the people would lose their identity".

13. Croatian:

Rekavši say-Rr	kako how-Cs	se self-Px-sa	"sjeća, remember-Vmr3s	te and-Cc	kako how-Cs
je be-Var3s	svjestan, aware-Agpmnsn	dubine depth-Ncfsg	(...)"	Peres speaker	je be-Var3s
izrazio express-Vmp-sm	sućut condolence-Ncfsg	(...)			

English: Saying he "remembers and aware of the depth (...)," Peres extended condolences (...)

Cross-branching

rough translation:
have described = su opisali

described

have

16. Croatian:		
Vojnici soldier-Ncmpn	su be-Var3p aux. part of the P	svoje their-own-Px-nsa
angažiranje engagement-Ncnsa	u in-Sl	Iraku Iraq-Npmsl
opisali describe-Vmp-pm participle part of the P	kao as-Cs	"QUOTE" quote

English: The soldiers described the engagement in Iraq as "QUOTE"

Passive

17. Croatian:

“QUOTE”,
quote

navodi se u priopćenju.
state-Vmr3s self-Px-sa in- Sl statement-Ncns
p a s s i v e

no speaker!

English: “QUOTE”, it is stated in the statement.

Sentence parts differently annotated

18. Croatian:

"Vjernik	sam	i	ponosim	se
believer-Ncmsn	be-Var1s	and-Cc	proud- Vmr1s	self-Px-sa
p o s i t i v e				
time,	ali	ubijanje	nevinih	ljudi
it-Pd-nsi	but-Cc	killing-Ncmsn	innocent-Agpmppy	people-Ncmpg
c o n t r a y c o n j u n c t i o n				
ne	može	nikako	biti	(...)"
no-Qz	can-Vmr3s	no way-Rgp	be-Van	
n e g a t i v e				

English: "I am faithful and proud of it, & but killing innocent people can not have (...)"

Conclusions

- we created a dataset with annotations from a corpus of 140 documents
- annotations consist of direct quotes and some direct parts of mixed quotes, their attributed speakers and verb-cues, along with their relations and sentence sentiment

Future work

- project the annotations on other languages in the SETimes corpus
- build the tool for automatic extraction, attribution and sentiment analysis of quotes



Thanks!

Questions

Acknowledgements:

The work presented has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 812997 and under the name CLEOPATRA (Crosslingual Event-centric Open Analytics Research Academy).

Credits

Presentation template by [Slidesgo](#)

Icons by [Flaticon](#), [Smashicon](#), [Darius Dan](#)

Images & infographics by [Freepik](#)

Author introduction slide photo created by **katemangostar** - Freepik.com

Big image slide photo created by **jcomp** - Freepik.com

Text & Image slide photo created by **rawpixel.com** - Freepik.com

Text & Image slide photo created by **Freepik**