

The figure simile

- Simile: an explicit comparison of something unknown or less known with something known
- Used either for praise or censure; the property described with the simile is intensified
- Form: ADJ as/like NOUN, VERB as/like NOUN

Our work in a nutshell

- Draws on research on Serbian similes and their encoding (Mitrović et al, 2015)
- Corpus and crowdsourcing based investigation of Greek similes
- Two lists of verified similes of the type “Adjective as Noun”
- Preliminary linguistic comparison of Serbian and Greek “adjective as noun” structures and of some “verb as noun” structures related to them.

Crowdsourcing Experiment

- **Aim:** obtaining a list of verified similes of the contemporary, “live” Greek language
- **Simile type:** “adjective as noun”
- **Background:** method tested on Serbian similes (Mladenović et al, 2016)
- **Method** applied on Serbian and Greek in independent experiments:
 - Potential similes were retrieved from corpora
 - Verification of the similes status as independent linguistic entities with a crowdsourcing experiment: native speakers marked the similes they use in their everyday usage of the language Google Forms were circulated via Facebook. Speakers clicked on constructs they used in their everyday exchange (Table 1).

SR “adjective as noun” similes:

- 2000 phrases with potential SR similes from the Corpus of Contemporary Serbian Language

GR “adjective as noun” similes:

- 2000 phrases with potential GR similes from the Hellenic National Corpus (HNC)
- ~500 phrases from a corpus (~110M words) obtained with web crawling

SERBIAN (SR)		GREEK (EL)	
Constructs per form	Participants per form	Constructs per form	Participants per form
30	46	30	67
42	138	42	85
41	150	41	79
41	100	41	59
154	434	154	290

TABLE 1. Distribution of constructs and participants in Greek CS project

Inter-annotator agreement checked with the Kappa Coefficient (Krippendorff 2004) (Table 2). 53/153 tested Serbian constructs elicited an inter-annotator agreement of 0,68-0,75 while 94/153 Greek constructs elicited an inter-annotator agreement between 0,65-0,8 (Table 2).

SERBIAN (SR)					GREEK (EL)				
Form set	Participants	Questions	“Yes” answers	Kappa value	Form set	Participants	Questions	“Yes” answers	Kappa value
1	5	30	16	0,7575	1	5	30	20	1
2a	5	21	17	0,713	2a	5	21	11	0,736
2b	5	21	15	0,698	2b	5	21	13	0,69
3a	5	21	5	0,688	3a	5	21	10	0,735
3b	5	20		0,484	3b	5	20	19	0,696
4a	5	21		0,434	4a	5	21	12	0,697
4b	5	19		0,375	4b	6	19	9	0,698
Totals	153	53			153	94			

TABLE 2. Results of the crowdsourcing experiment for SR and EL similes

Do similes have the status of a MWE?

Similes function as the superlative of the adjective that would denote the property they denote.

• Varying semantic transparency:

EL άσπρος σαν το πανί **SR** beo (bled) kao krpa –
white as the cloth as white as kitchen cloth
'very pale, sickly looking'

• Exert strict selection restrictions on the nouns they are predicated of

EL άσπρος σαν το πανί selects humans or body parts (face, cheeks, lips)
EL στολισμένος σαν φρεγάτα (“accessorized like frigate”) strongly prefers +human, +feminine (in a Google search, it returned 13 unique hits of which only 1 was +masculine)
SR bled kao krpa (“white (pale) as the cloth”) prefers +human, +feminine

• Form collocations with the nouns they are predicated of

EL απαλός σαν μετάξι (“soft as silk”) applies on skin BUT
απαλός σαν χάδι (“soft as stroke”) rarely selects skin
(Google: απαλός σαν μετάξι: ~210 unique results, >60 skin
απαλός σαν χάδι: ~225 unique results, only 3 skin)

• ADJ inclines fully and probably NOUN agrees in number with the ADJ.

EL γηραιές κυρίες στολισμένες σαν φρεγάτες
old.PL lady.PL accessorised.PL as frigate.PL

• not really flexible word order, internal modification possible but not frequent

The results of this study will be:

- encoded in IDION – an encoding environment for GR MWEs (Markantonatou et al. 2015)
- used to enrich the Ontology of Rhetorical Figures for Serbian (Mladenović and Mitrović, 2013) with Greek similes allowing us to perform and encode a deeper comparison of the two languages
- connected with a conceptually organised lexicon such as the Greek WordNet (Tufiş, Cristea and Stamou, 2004) – Serbian simile were already added to Serbian WordNet via new Cross-POS semantic relation specificOf/specifiedBy (Mladenović et al, 2016).

Preliminary comparison of SR /EL similes

1. Same concept is expressed by literal translational equivalents:

- ADJ like NOUN similes where ADJ in the semantic field of color (colors for which EL/SR similes were found: white, black, red, yellow)

EL άσπρο σαν το γάλα **SR** beo kao mleko – ‘very white’
white as the milk white as milk
EL άσπρο σαν το χιόνι **SR** beo kao sneg – ‘very white’
white as the snow white as snow
EL άσπρο σαν το πανί **SR** beo (bled) kao krpa
white as the cloth white as cloth

‘very pale – negative sentiment in both languages, meaning that someone is very pale, very sickly looking.

- ADJ like NOUN similes about characteristics attributed to animals by folk wisdom

EL πονηρός σαν αλεπού **SR** lukav kao lisica – ‘very sly’
sly as fox sly as fox
EL πιστός σαν σκύλος/σκυλί **SR** veran kao pas – ‘very faithful’
faithful as dog faithful as dog
EL βρώμικος σαν γουρούνι **SR** prljav kao svinja – ‘filthy’
dirty as pig dirty as pig

- Other shared ADJ like NOUN similes

EL λαμπρός σαν τον ήλιο **SR** Sjajan kao sunce – ‘very bright’
bright as the sun bright as sun
EL χοντρός σαν βαρέλι **SR** Debeo kao bure – ‘very fat’
fat as barrel fat as barrel

2. Same concept is expressed with different similes

Here are concepts that are expressed with similes in both languages but the similes are different. For example, the concept “something is easy” is expressed with a simile in both Serbian and Greek. The Serbian simile is *prosto kao pasulj*, literally, “As easy as (cooking) beans”. The Greek simile is the verb simile *κάτι είναι παιχνιδάκι* “something is a game” and there is no similar food related expression in Greek.

3. We have not encountered any similes that are word by word equivalents but denote different concepts.

Morphological relations across similes with heads of different PoS

– [deverbal ADJ] as NOUN similes can be related to VERB as NOUN similes
EL similes with a deverbal adjective very often have verb MWE counterparts with the base verb as a head verb and the σαν + NOUN as a complement.

EL ADJ γυρίζει από την αλάνα πεινασμένος σαν λύκος
returns from the sandlot hungry as wolf
'He is starving when he comes back from the sandlot.'

VERB μετά το θέατρο πεινάω σαν λύκος
after the theater feel hungry as wolf
'I starve after the theater'

– SR and EL de-adjectival verbs in the semantic field of colour appear in similes parallel to the similes formed with the base adjective.

SR crven kao bulka “red as a poppy”
pocrveneo kao bulka “blush as a poppy”

EL κόκκινος σαν παπαρούνα “red as a poppy”
κοκκινίζω σαν παπαρούνα “blush as a poppy”

References

- Mladenović, Miljana, Jelena, Mitrović and Cvetana Krstev. 2016. A Language-independent Model for Adding New Semantic Relations to a WordNet. 8th Global WordNet Conference.
- Mitrović, Jelena, Miljana Mladenović and Cvetana Krstev. 2015. “Adding MWEs to Serbian Lexical Resources Using Crowdsourcing” 5th General PARSEME meeting, September 23rd, 2015.
- Markantonatou, Stella, Erasmia Koletti, Elpiniki Margariti, Panagiotis Minos, Aimilia Stripeli, Georgios Zakis and Niki Samaridi. 2015. Lexical Resource for free subject verb MWEs. 4th General Parseme Meeting, Malta, March 19-20, 2015.
- Utvić, Miloš. 2014. Corpus of Contemporary Serbian – Frequency Lists. Naučni sastanak slavista u Vukove dane, 241–262. Faculty of Philology, University of Belgrade, Belgrade.
- Mladenović, Miljana and Jelena Mitrović. 2013. Ontology of Rhetorical Figures for Serbian. Lecture Notes in Computer Science and Artificial Intelligence 8082. Springer-Verlag. Berlin Heilderberg. 386-393.
- Tufiş, Dan, Dan Cristea, and Sofia Stamou. 2004. BalkaNet: Aims, Methods, Results and Perspectives. A General Overview. 9-43
- Krippendorff, Klaus. 2004. Content Analysis: An Introduction to Its Methodology. Sage.