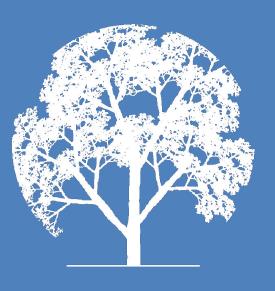


2<sup>nd</sup> General Meeting University of Naples "L'Orientale" Naples, Italy, 8-9 February 2024

https://unidive.lisn.upsaclay.fr/



# **Treating Multiword Expressions with a view** to Morphologically Rich Languages



Svetlozara Leseva | zarka@dcl.bas.bg | https://dcl.bas.bg/ Department of Computational Linguistics Institute for Bulgarian Language, Bulgarian Academy of Sciences







## Objective

- A uniform linguistic description focusing on:
  - the representation of the structural, morphological, morphosyntactic, word-order, and other features of Bulgarian MWEs;

### A lexicon of MWEs

- The lexicon includes:
  - over 10,000 nominal MWEs
- extension to MWEs for other languages;
- o with a view to the automatic recognition and annotation of MWEs in running text.

A layered linguistic description

Modularity: the inflection type of each MWE is defined as a sequence of elementary fields describing the individual MWE components in a way that allows us to model independently them and thus the MWEs':

- lexicogrammatical characteristics;
- inflectional morphosyntactic features: characteristics, morphosyntactic constraints;

- including 5,000 NEs
- o 6,500 verbal MWEs:

пимфен възел

- 1,200 light verb constructions;
- 1,800 verbal idioms;
- 3,400 reflexives and others.

# Validation of MWE forms in BuINC

• The MWE forms are automatically generated and: validated against the Bulgarian National Corpus (the attested forms are marked by a  $\checkmark$ ); OR heuristically confirmed (not marked); OR non-confirmed (in gray):

informatsionni tehnologii (pluralia tantum) 'information technologies' = plural

(Az) imam zlatno sartse – (Nie) imame zlatni sartsa

(direct object agreeing in number with the subject)

'(I) have a heart of gold' - '(We) have hearts of gold'

- structural characteristics:
  - internal syntactic structure;
  - the list of head and dependents;
  - possible variations in their linear order.
- syntagmatic properties:
  - syntactic transformations and constraints;
  - optional components and possible insertions:

vzemam reshenie – vzemam **vazhno** reshenie

<ul> <li>лимфен възел:1; лимфна жлеза:1; всяко едно от малките телца, изградени от лимфна тъкан, разположени на интервали по хода на лимфните съдове; откриват се в много части по тялото, например в слабините, в аксилата, зад ухото; действат като филтри за лимфата, като не позволяват на чуждите частици да попаднат в кръвообращението, произвеждат и лимфоцити</li> </ul>
<ul> <li>лимфният възел NMsl</li> <li>лимфния възела NMsh</li> <li>лимфни възела NMpb</li> <li>лимфни възли NMpo</li> </ul>
<ul> <li>лимфните възли NMpd ✓</li> <li>лимфна система Мед. NHF11 фикс. eng-30-05396366-n –</li> </ul>
<ul> <li>лимфна система NFso </li> <li>лимфната система NFsd </li> <li>лимфни системи NFpo</li> <li>лимфните системи NFpd</li> </ul>

Nominal MWEs (limfen vazel 'lymph node', limfna sistema 'lymphatic system') with morphosyntactic and semantic description, link to WN synset and validation in BulNC.

# Description of verbal MWEs

Synset ID / MWE ID	eng-30-02524739-v / bg_2291
MWE lemma / Abstract lemma	удрям джакпота / удрям джакпот
Morphosyntactic features	удрям.V_IMPERF_r1s джакпота.Nsh
Head and head inflection type	<i>удрям</i> .V_IM_TT_S3_01
Head restrictions	none
Dependent and dep restrictions	джакпота / fixed; <b>N</b> (umber) = s; <b>D</b> (efiniteness) = h
Syntactic structure	Constituent: V N(P)   UD: V + obj
Semantic frame	Success_or_failure (Agent, Goal   Role)
Subcategorisation: subject	N(P)_subj   UD: nsubj
Subcategorisation: complements	Goal: PP   UD: obl & P = в/във; Role: PP   UD: obl & P = като
Possible modifiers of the head	regular
Possible modifiers of dependent	regular; A(P); Ex.: удрям <b>големия/Ash</b> джакпот/Ns0
External elements	regular (question particle   subj   AdvP)
PARSEME type	VID
Register and connotation	Colloquial; -0.125 +0.25
Derivational relations	удряне на джакпота

### 'make a decision' – 'make an **important** decision'

possible derivations and other transformations:

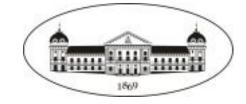
vzemam reshenie – vzemane na reshenie 'make a decision' – '(the) making of a decision'

• (if applicable): subcategorisation frame, argument transformations and selectional restrictions.

Description of verbal wives.

\*In pink: directions for future work.

#### Acknowledgements:



This research is carried out as part of the project Semantic Resources and Language Technologies (Lexical-Semantic Networks and Language) Models) of the Institute for Bulgarian Language, Bulgarian Academy of Sciences (2023–2025).