

# A Valency Lexicon for Universal Dependencies

**Petr Kocharov, Lilit Kharatyan**

petr.kocharov@uni-wuerzburg.de

lilith.kharatian@gmail.com

Julius-Maximilians-Universität Würzburg

## What is UDVal: Valency Lexicon for Universal Dependencies?

---

- A prototype of a web-application for building valency lexica from treebanks
  - stored in the CoNLL-U format and following the UD annotation scheme  
<https://universaldependencies.org>
- Derived from the Classical Armenian Valency Lexicon  
<https://caval.dch.phil-fak.uni-koeln.de>
  - by differentiating the core treebank-independent functionality and treebank-specific extensions
  - the core functionality can be reused for any validated UD treebank thanks to highly standardized UD annotation conventions

## Functionality and applications

---

- Automatically retrieves verb frames for all verbs in a treebank
- Supports queries based on verbs and parameters of their dependents
  - syntactic relations
  - encodings
  - lemmas
- Frequency-informed research on
  - valency classes
  - encoding of arguments and adjuncts
  - valency alternations and morphosyntactic variation patterns
  - alignment (flagging, indexing)

## What is UDVal for?

---

Case studies in Classical Armenian highlighting usability of the tool

- frequency data on split S, A, P, and R coding
- establishing the basic transitive and ditransitive alignment
- lexical and morphological features of transitivity pairs (cf. “World Atlas of Transitivity Pairs”: <https://watp.ninjal.ac.jp/en/>)
- valency classes and valency alternations for the “Pavia Verbs Database” (<https://paveda.unipv.it/contributions/clas1256#tverbs>)

## Why UDVal?

---

- General-purpose tools to query UD treebanks often
  - require knowledge of formal query languages (PML-Tree Query; TüNDRA)
  - are not provided with an online interface (UDEasy)
  - not customized for research on verb frames
- UDVal offers a flexible user-friendly online interface that significantly facilitates access to corpus evidence on verb frames provided with standardized typologically oriented morphosyntactic annotation

## Documented codes ready for reuse

- [https://github.com/caval-project/ud\\_val](https://github.com/caval-project/ud_val)



GitHub repository view for `caval-project / ud_val`. The interface shows navigation tabs for Code, Issues, Pull requests, Actions, Projects, Security (79), Insights, and Settings. The left sidebar displays the file tree for the `app` directory, including `static`, `templates`, `extensions.py`, `init.py`, `query_helpers.py`, `routes_language.py`, `routes_translit.py`, `README.md`, and `requirements.txt`.

The main content area shows the `ud_val / app /` directory view, featuring a table of files and folders with their last commit messages:

Name	Last commit message
..	
static	Add files via upload
templates	Create translit.html
extensions.py	Create extensions.py
init.py	Create init.py
query_helpers.py	Create query_helpers.py
routes_language.py	Create routes_language.py
routes_translit.py	Create routes_translit.py

## App architecture

---

- Backend
  - MariaDB relational database
  - automatic population from CoNLL-U
  - parameterized SQL queries
  - indexing and query optimization
- Application layer
  - Python + Flask
  - Dynamic query processing
- Frontend
  - HTML, CSS, JavaScript
  - responsive search interface

Pipeline: UD treebank → database → query engine → verb lists → frame visualization

## Verb frames with UD

---

UDVaL stores a verb together with types of dependents specified below

- Nominal
  - `nsubj`, `obj`, `iobj`, `obl`
  - `nmod` (for nominalized heads)
- Clausal
  - `csubj`, `ccomp`, `xcomp`
- Auxiliary
  - `aux` (for analytic verb forms)
- Nominal dependents are stored together with their encoding (Case + adposition), following the UD's treatment of adpositions on a par with case marking.

## Classical Armenian Valency Lexicon

[Search](#) [About](#) [How to Use](#) [Publications](#)

### Select texts

Select texts

### Search by verb

Search verb

Search meaning

### Search by features

Show verb features

### Search by valency

Dependency 1

obj

Select encoding

Select lemma

Dependency 2

iobj

Select encoding

Select lemma

Remove

Add

**a b g d e z t' x c h m y n š o p s t c' k' | All**

Orthography: [Armenian script](#)

Number of verbs: 59

Number of occurrences: 494

Sort: [alphabetically](#) ▼ / [by frequency](#)

acel 'drive' (1)	erewec'owc'anel 'imagine' (1)	patmel 'tell' (10)
agowc'anel 'make put on' (1)	ganjel 'deposit' (2)	patowirel 'order' (3)
aṙak'el 'send' (3)	gnel 'buy' (2)	patrastel 'prepare' (5)
arbowc'anel 'make drink' (2)	gtanel 'find' (1)	šinel 'build' (1)
arjakel 'release' (14)	hanel 'take away' (2)	sirel 'love' (1)
arkanel 'throw' (2)	harc'anel 'ask' (6)	šnorhel 'grant' (1)
arnel 'make' (30)	harkanel 'strike' (2)	tal 'give' (225)
asel 'say' (23)	hatowc'anel 'compensate' (2)	tanel 'carry' (1)
awetaranel 'proclaim the gospel' (3)	hawatal 'believe' (2)	t'anal 'wet' (1)
banal 'open' (1)	k'arozel 'preach' (2)	t'awalec'owc'anel 'roll down' (1)
baṙnal 'raise' (1)	matnel 'betray' (13)	t'ołowl 'let' (12)
bašxel 'distribute' (1)	matowc'anel 'offer' (7)	xawsel 'speak' (4)
bažanel 'divide' (1)	meknel 'declare' (3)	xndrel 'ask' (1)
berel 'bring' (2)	melanč'el 'sin' (1)	yarowc'anel 'raise' (5)
canowc'anel 'announce' (2)	nmanec'owc'anel 'compare' (5)	yawelowl 'increase' (1)
cnanel 'be born/give birth' (1)	ownel 'have' (18)	yaytnel 'reveal' (5)
c'owc'anel 'show' (22)	owsowc'anel 'teach' (2)	yišec'owc'anel 'remind' (1)
darjowc'anel 'turn' (1)	owxtel 'vow' (1)	zenowl 'sacrifice' (1)
dnel 'put' (7)	paganel 'worship' (19)	zgec'owc'anel 'wear' (3)
erdnowl 'swear' (1)	partel 'defeat' (2)	

## Search interface and interaction

---

Users can search by any combination of the following parameters

- verb lemma
- dependent relation
- dependent encoding
- dependent lemma

Core interaction model

- set parameters for the first dependent
- add/remove additional dependent with its own set of parameters
- queries can be refined from simple to highly specific verb-frame patterns
- exhaustive lists of verbs or their occurrences are dynamically extracted from the corpus and visualized in the output area of the webpage

## Interface features

---

Users gets:

- filtered lists of verbs with frequency counts
- frequency-based or alphabetical ordering of verb lists
- filtered lists of verb occurrences with frequency counts
- BRAT-based visualization of filtered verb frames in occurrences
- grammatical glossing on mouse-over
- reference to treebank sentence IDs

Additional strengths:

- dynamic frequency feedback
- immediate evidence for frame availability
- usable for both single verbs and verb classes

## Classical Armenian Valency Lexicon

[Search](#) [About](#) [How to Use](#) [Publications](#)

### Select texts

Select texts -

### Search by verb

Search verb

Search meaning

### Search by features

Show verb features -

### Search by valency

Dependency 1

obj

Acc + z

Select lemma

Dependency 2

iobj

Dat

Select lemma

[Remove](#) [Add](#)

[a](#) [b](#) [d](#) [e](#) [z](#) [t'](#) [x](#) [c](#) [h](#) [m](#) [y](#) [ñ](#) [o](#) [p](#) [t](#) [c'](#) [k'](#) | [All](#)

Orthography: [Armenian script](#)

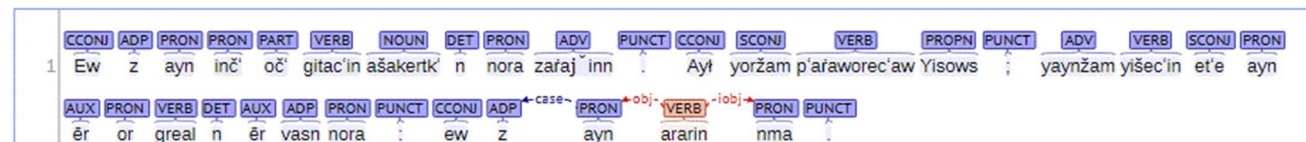
Number of occurrences: 6

[aṙnel 'make'](#) [↗](#)

Sentences 1–6

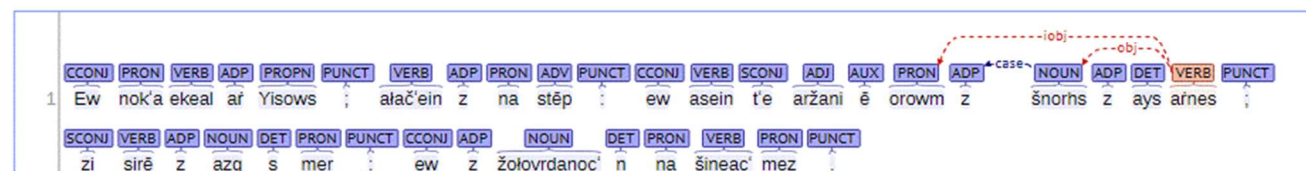
Gospel of John 12:16 (sentence\_id: JOHN\_12.16):

Ew zayn inč' oč' gitac'in ašakertk' n nora zaṙaj'inn. Ayt' yoržam p'araworec'aw Yisows; yaynžam yišec'in et'e ayn ēr or grealn ēr vasn nora: ew zayn ararin nma. [↗](#)



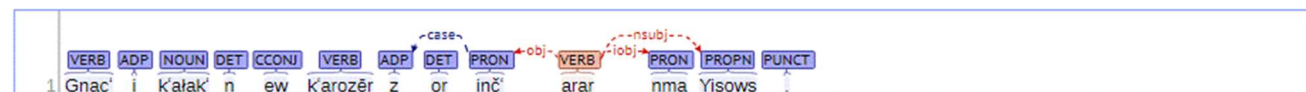
Gospel of Luke 7:4-5 (sentence\_id: LUKE\_7.4-5):

Ew nok'a ekeal aṙ Yisows; alač'ein zna stēp: ew asejn t'e aržani ē orowm z šnorhs z ays arnes; zi sirē zazgs mer: ew zžoovrdanoc'n na šineac' mez. [↗](#)



Gospel of Luke 8:39 (sentence\_id: LUKE\_8.39b):

Gnac' i K'alak' n ew K'arozēr z or inč' arar nma Yisows. [↗](#)



## Potential functional extensions

---

Extensions already implemented for the Classical Armenian treebank (reusable for all UD treebanks with shared properties)

- Latinized input
- Transliteration
- Search by English semantic gloss
- Sentence translation display
- Links to external lexicographic resources
- Sub-corpus selector
- Filtering by morphological features of verbs

## Limitations and Future Work

---

### Current limitations

- some extensions still require manual adaptation
- currently focused on verbal heads and a restricted set of dependents
- no unified interface for cross-treebank comparison
- response time increases with larger corpora
- output quality depends on the quality of underlying treebank annotation

### Future directions

- automatic detection of treebank-specific features
- broader coverage of predicate types and constructions
- support for multi-treebank comparison
- improved scalability and response time

## Conclusion

---

UDVaL converts UD treebanks into searchable online valency lexica combining

- typologically informed design
- user-friendly querying
- corpus-based linguistic evidence

It is especially useful for researchers who need fast and precise access to verb frame data.

# A Valency Lexicon for Universal Dependencies

**Petr Kocharov, Lilit Kharatyan**

petr.kocharov@uni-wuerzburg.de

lilith.kharatian@gmail.com

Julius-Maximilians-Universität Würzburg