

## **NOTA BENE:**

În acest document, expresiile folosite pentru interogarea CoRoLa sunt scrise, pentru distingerea lor de restul textului, cu fontul Courier New și pe rând separat.

**Acest document nu înlocuiește documentația KorAP existentă pe site, ci oferă doar câteva exemple pentru limba română. Utilizatorul este încurajat să folosească documentația KorAP (<http://89.38.230.10:5555/doc/korap>).**

**Acest document conține două părți:**

- 1. Interogarea CoRoLa cu KorAP** – aceasta prezintă diverse moduri în care utilizatorul își poate concepe fraza de interogare a corpusului, având la dispoziție, pe site, fișierele cu informații despre procedura de adnotare;
- 2. Interogarea bazată pe exemple a corpusului CoRoLa cu ajutorul KorAP** – aceasta prezintă modul în care utilizatorul își poate concepe fraza de interogare folosindu-se de analiza unor exemple deja găsite în corpus.

## **1. Interogarea CoRoLa cu KorAP**

Este permisă atât utilizarea cuvintelor, cât și a expresiilor regulate în căutări, după cum exemplificăm mai jos.

### **CĂUTAREA DE TIP *LEXICAL***

Căutarea unor **cuvinte simple (în forma în care apar în corpus)** se poate face prin introducerea lor în bara de căutare:

`copil`

SAU folosind **atributul „orth”**:

`[drukola/orth=copil]`

Cuvântul va fi căutat cu tipul de litere introduse (în exemplele de mai sus, mici).

Căutarea unui **cuvânt indiferent de tipul de litere (mici/mari)** cu care este scris se face adăugând /i după cuvântul respectiv:

`copil/i`

SAU:


`[drukola/orth=copil/i]`

**Frecvența** cuvântului/cuvintelor căutat/e este afișată în dreapta, deasupra primului rezultat, **DACĂ NU ESTE BIFATĂ** caseta Glimpse.

Rezultatele căutării nu se pot **sorta** și nici **grupa**, deocamdată. Aceste opțiuni **vor fi** disponibile în viitorul apropiat.

Coloana din stânga afișează calea și numele documentelor în care au fost găsite rezultatele.

Un click pe rezultat permite afișarea unui context mai larg, precum și a titlului, autorului și anului de apariție ale fiecărui text (pe fond portocaliu, în stânga; în partea dreapta rămâne numele fișierului).

Pentru afișarea formei lematizate și adnotate morfosintatic a propoziției în care a fost găsit cuvântul căutat se va face click pe simbolul 

Căutarea după **lemă (folosind atributul "base")**:

```
[drukola/base=copil]
```

Căutarea folosind operatori logici:

- ex.: căutarea formei „acele” ca substantiv

```
[drukola/orth=acele & drukola/m="msd:n.*"]
```

**Atributul „m”** indică o căutare la nivel morfologic după eticheta MSD (vezi mai jos). Expresia regulată “n.\*” selectează toate etichetele MSD care încep cu litera n (sunt substantive) și sunt urmate de 0, 1, sau mai multe (\* închiderea Kleene) caractere oarecare (simbolul “.” stă pentru orice caracter).

SAU:

```
[drukola/orth=acele & drukola/base=ac]
```

A se vedea diferența între această cerere și următoarea:

```
[drukola/orth=acele]
```

Numai aparent redundantă, condiția `drukola/base=ac` elimină ocurențele cuvântului „acele” ca determinant sau pronume.

- toate ocurențele substantivului „tată” care NU sunt în vocativ:

```
[drukola/orth=tată & drukola/m!=case:vocative]
```

Căutarea a două (sau mai multe) **cuvinte aflate în vecinătate imediată** în text (! Pot apărea semne de punctuație între ele.):

`copil cuminte`

SAU:

```
[drukola/orth=copil] [drukola/orth=cuminte]
```

Căutarea fără și cu **negație**:

```
[drukola/base=putea] [drukola/base=să]
```

```
[drukola/base=putea] [drukola/base!=să]
```

Căutarea secvențelor cu **repetare**: de exemplu: o succesiune de trei determinatori:

```
[drukola/m="msd:d.*"] {3}
```

Căutarea cuvintelor **coocurente aflate la distanță unul față de altul**: folosirea simbolului pentru *segmentul vid* (`[]`) care regăsește orice cuvânt (nu și punctuație!). Exemple:

- la distanță de un cuvânt:

```
[drukola/base=copil] [] [drukola/base=cuminte]
```

- la distanță de cel puțin două cuvinte:

```
[drukola/base=copil] []{2, } [drukola/base=cuminte]
```

- la distanță de cel puțin două cuvinte și cel mult cinci cuvinte:

```
[drukola/base=copil] []{2,5} [drukola/base=cuminte]
```

**Căutarea într-un subcorpus**. Atunci când nu se precizează de către utilizator în ce fel de texte să se caute, KorAP va căuta în TOT corpusul indexat. Dacă utilizatorul dorește să își restrângă căutarea la textele dintr-un anumit stil funcțional, apărute într-un anumit an etc., poate face acest lucru parcurgând următorii pași:

- click pe „all Corpora”, în stânga, sub bara de căutare;

- click pe punctele de suspensie ce au apărut deasupra „all Corpora”;

- selectarea (prin click) din listă a elementului de metadate (atributului) cu ajutorul căruia se dorește restrângerea căutării;

- selectarea relației de interes: prin click pe **eq** apar toate variantele: eq (echivalează cu), ne (nu echivalează cu), contains (conține), containsnot (nu conține)
- introducerea valorii/valorilor de interes pentru acel element (atribut) prin click pe punctele de suspensie.

Restrângerea corpusului de lucru se poate face chiar mai mult, prin apăsarea operatorilor logici „and”, „or” sau se poate elimina prin apăsarea x-ului.

Căutarea de **cuvinte care conțin un anumit șir de litere:**

-oriunde în cuvânt:

"hiper"/x

- la începutul cuvântului: ex.: cuvinte cu **prefixoidul hiper-**

"hiper.+"

- la mijlocul cuvântului: ex.: cuvinte care conțin **triftongul oai:**

".+oai.+"

- la finalul cuvântului: ex.: căutarea de cuvinte **sufixate** cu **-tor:**

".+tor"

## CĂUTAREA DE TIP MORFOLOGIC

Căutarea după elemente morfologice se poate face utilizând codificările descrierilor morfosintactice MULTEXT-East (MSD-uri) sau trăsături sub forma perechilor atribut-valoare.

Exemple de căutări de tip morfologic:

- verbe la **participiu:** expresia de căutare cu MSD-uri:

[drukola/m=msd:vmp--sm]

SAU cu trăsături:

[drukola/m=verbform:participle]

- verbe la **supin:** expresia de căutare cu MSD-uri (Introduceți ghilimelele cu atenție în interfața de căutare: nu toate tipurile de ghilimele sunt recunoscute în formularea expresiei regulate!):

[drukola/m="msd:s.\*"] [drukola/m="msd:vmp.\*"]

SAU cu trăsături:

[drukola/m=pos:adposition] [drukola/m=verbform:participle]

- secvențe **substantiv sau adjectiv + prepoziție + substantiv:** expresia de căutare cu MSD-uri :

[drukola/m="msd:n.\*" | drukola/m="msd:a.\*"] [drukola/m="msd:s.\*"]

[drukola/m="msd:n.\*"]

SAU cu trăsături:

[drukola/m=pos:noun] [drukola/m=pos:adposition]

[drukola/m=pos:noun]

## COMBINAREA NIVELURILOR ÎN CĂUTĂRI

Exemplu: căutarea unui cuvânt după **lemă + morfologie:**

[drukola/base=președinte & drukola/m=case:oblique &

drukola/m=number:plural]

## 2. Interogarea bazată pe exemple a corpusului CoRoLa cu ajutorul KorAP

La vizualizarea adnotării morfosintactice pentru un exemplu, se pot selecta caracteristici care să constituie componentele unei noi interogări. În exemplul de mai jos observăm cum au fost selectate

(devenind marcate cu verde) două caracteristici pentru o nouă căutare: un cuvânt cu atributul pos având valoare nsry (substantiv comun, singular, la un caz direct, articulat hotărât) urmat la distanță de două cuvinte de un alt cuvânt având atributul msd cu valoarea vmip3 (verb predicativ, modul indicativ, timpul prezent, persoana a treia).

The screenshot shows the KorAP search interface with the query `[drukola/m=pos:nsry][2][drukola/m=msd:vmip3]`. The search results are displayed in a table format. The table has columns for Foundry, Layer, and various morphological and syntactic features. The results are filtered to show only those with the specified attributes.

Foundry	Layer	Odorul	acesta	se	păstrează	în	muzeul	din
drukola	l	odor	acesta	sine	păstra	în	muzeu	din
drukola	m	case:direct definiteness:yes gender:masc msd:ncmsry number:singular pos:nsry type:common	case:direct gender:masc modifiertype:postnominal msd:dd3msr--o number:singular pos:dmsr type:demonstrative	case:accusative msd:px3-a-----w person:third pos:pxa pronform:weak type:reflexive	msd:vmip3 person:third pos:v3 tense:present type:main verbform:indicative	case:accusative msd:spsa pos:s type:preposition wordformation:simple	case:direct definiteness:yes gender:masc msd:ncmsry number:singular type:common	case:accusative msd:spsa pos:s type:preposition wordformation:simple
drukola	p	noun	determiner	pronoun	verb	adposition	noun	adposition

At the bottom of the table, there is a "New Query" bar containing the query `[drukola/m=pos:nsry][2][drukola/m=msd:vmip3]`.

Făcând clic pe bara gri de jos, care începe cu „New Query” și conține această nouă frază de interogare, se vor căuta în corpus exemple și se vor întoarce rezultate după cum exemplificăm mai jos:

The screenshot shows the KorAP search interface with the query `[drukola/m=pos:nsry][2][drukola/m=msd:vmip3]`. The search results are displayed in a list format. The results are filtered to show only those with the specified attributes.

Corola-other/Imaginative/83522\_a\_84852 Ștefan cel Mare. **Odorul acesta se păstrează** în muzeul din București.

Corola-other/Imaginative/83522\_a\_84847 .f.131) Zice că-n **locul care se cheamă** Negriștei, din comuna Doagele, ținutul Vaslui, ar fi fost în vremea lui Ștefan cel Mare un târgușor. Numele târgușorului

Corola-other/Imaginative/83530\_a\_84855 Bărcălabu a fost în **voia ta, acum lasă** să fie în voia mea. Mai pe urmă și-a tras Vodă cu seamă, și l-a iertat de copilăria care-o făcuse, după cum l-a iertat și pe b

Corola-other/Imaginative/83496\_a\_84821 valea Similei, pe **dealul care se cheamă** Cetatea, unde se văd și astăzi 2 tabii (șanțuri); pe urmă la gura văii Racova și de-acolo s-a tot bătut în sus pe apa Bă

Corola-other/Imaginative/83494\_a\_84819 ce. Înțelegerea și **definirea lor se structurează** într-un proces diversificat și construit pe mai multe niveluri. Tradiția istoriei rămâne, în esență, o relație rezu

Corola-other/Imaginative/83536\_a\_84851 nuce, și nu știa că **luarea Griviței nu înseamnă** luarea Plevnei, și că va sta de strajă multe nopți reci și lungi de-a rândul și că-i vor amorți de frig picioarele-n sa

Corola-other/Imaginative/83884\_a\_85209 ezii care acoperă **partea carosabilă, mașinile derapează** cu ușurință, sporind dificultatea conducerii și opririi lor în siguranță, odată cu posibilitatea comiterii

Corola-other/Imaginative/83884\_a\_85209 ația artistică de la **școala noastră vă prezintă** programul „Stop la neatenție ! ” cu aspecte legate de circulația rutieră. ELEVUL „A” : Circulația corectă E proble

Corola-other/Imaginative/83884\_a\_85209 mă ? Să traversezi **strada ! Așa te îndeamnă** . Traversez cu grijă Fără nici un zor și privesc lumina Către semafor. Semaforul dară Ni-i prieten bunCulorilor sale V

Corola-other/Imaginative/83884\_a\_85209 ben, nu pornești. **Verdele dacă te cheamă** Tu să treci fără de teamă. Semaforul spune : „Hai ! Acum treci ! Acum mai stai !” CONDUITA PREVENTIVĂ La ora d

Corola-other/Imaginative/83884\_a\_85209 ărtat suficient de **locul respectiv. Nu trebuie** să avem încredere că pe „trecerea de pietoni” șoferii ne feresc întotdeauna. Există momente când nici ei nu mai

Corola-other/Imaginative/83884\_a\_85209 hiculelor, sporind **riscul de accident. Trebuie** găsită o curte mai mare, un câmp, o alee, un teren de sport etc., unde nu circulă autovehicule. Nu uita că, odată

Corola-other/Imaginative/83884\_a\_85209 numele străzii și **numărul casei și întreabă** un agent de circulație cum se ajunge la acea adresă. - Când merge cu colegii, Gabi se amuză grozav punându-le p

Corola-other/Imaginative/83884\_a\_85209 ă el ne vrea doar **binele și ne apără** de accidente. Este cu adevărat prietenul copiilor. AM O BICICLETĂ NOUĂ Mare mi-a fost bucuria când

Corola-other/Imaginative/83884\_a\_85209 e face orizont. În **felul acesta El Uniformizează** pre-jurul. Dar fără prefacere, Fără neantizare, E nevinovat. El fâurește pământ Lincezind luciul subțire Tăietor

Observăm cum în bara de căutare avem fraza de interogare pe care am creat-o mai devreme, folosindu-ne de adnotarea exemplilor regăsite într-o altă căutare.