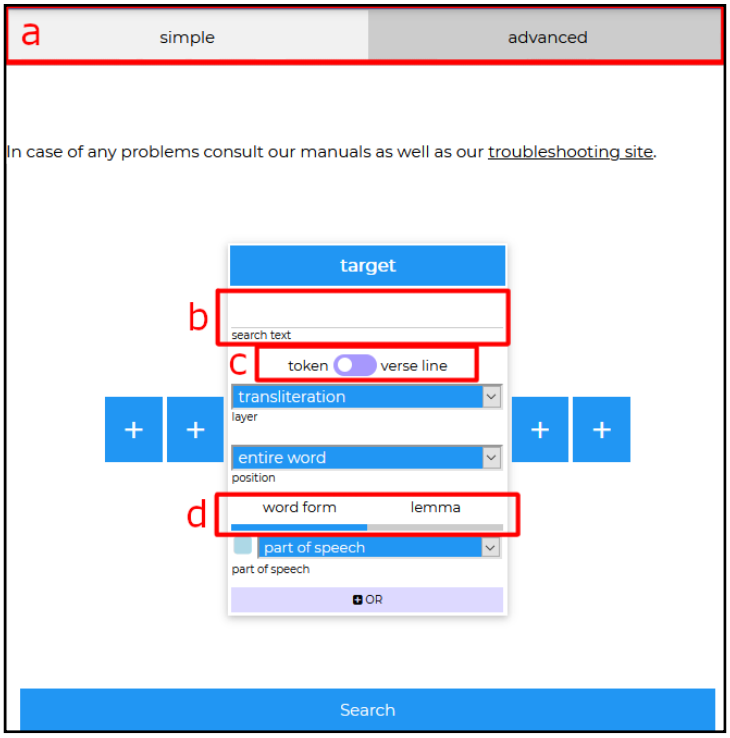
CorDon – The Advanced Search

CorDon's search mask has two modes, **simple** and **advanced**. The buttons at the top (a, img. 1) switch between the two. The functions of the simple search are explained here.



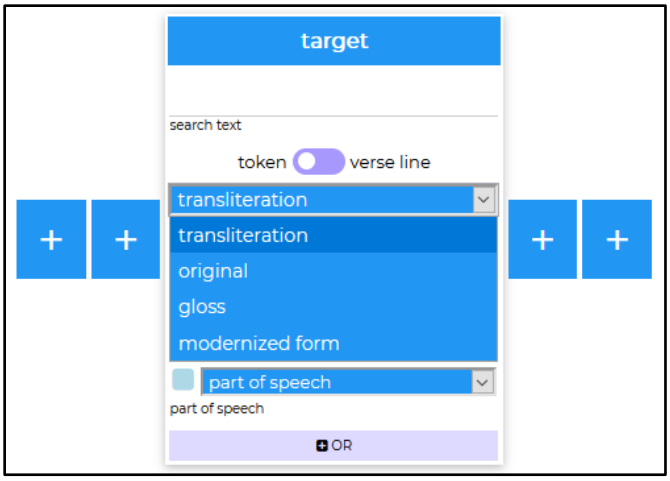
Img. 1

The advanced search mask is structured so that the search can be **narrowed step by step**. Img. 1 shows the view at the start.
The search text is entered at b. The token–verse line switch (c) switches the search between **token mode** and **verse line mode**.
As in the simple search, token mode can search either on the level of concrete **word forms** or **lemmata** (citation forms, "dictionary" forms) (d).

Note that when the search is set to **Lemma**, it doesn't matter which text layer (cf. img. 2) is selected.

The token search

Layer



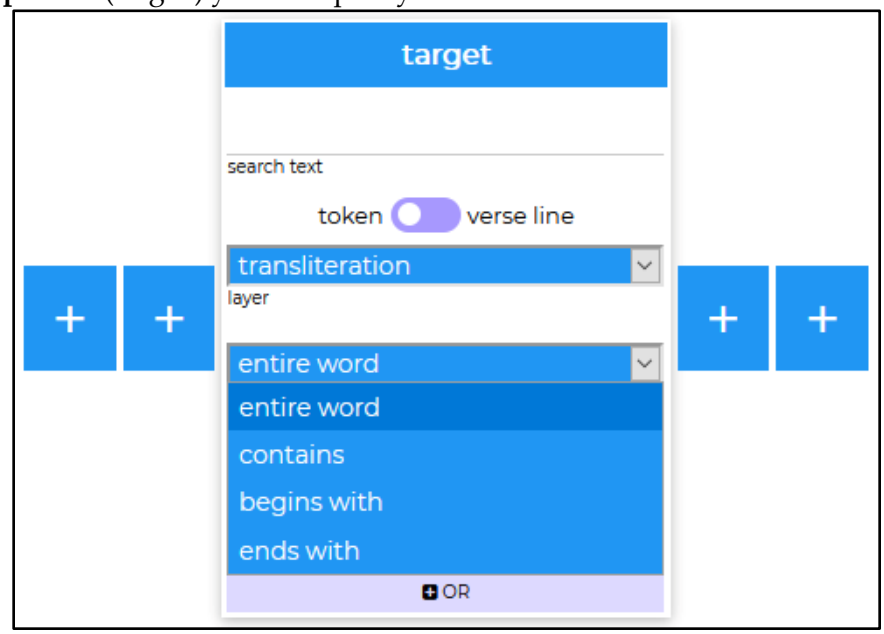
Img. 2

From the **"layer" dropdown** (img. 2) you can pick the text layer to be searched. By default, as in the simple search, the **transliterated** layer will be searched. Here the text has been rendered in the modern Lithuanian alphabet but left otherwise unchanged (cf. the description of the simple search). Alternatively the search can be set to:

- The **original** text, containing all special characters, metrical diacritics etc.
- The **glosses**, where obsolete, ambiguous or otherwise "difficult" words are glossed
- The **modernized form** – here, all word forms have been represented by their modern standard Lithuanian equivalent

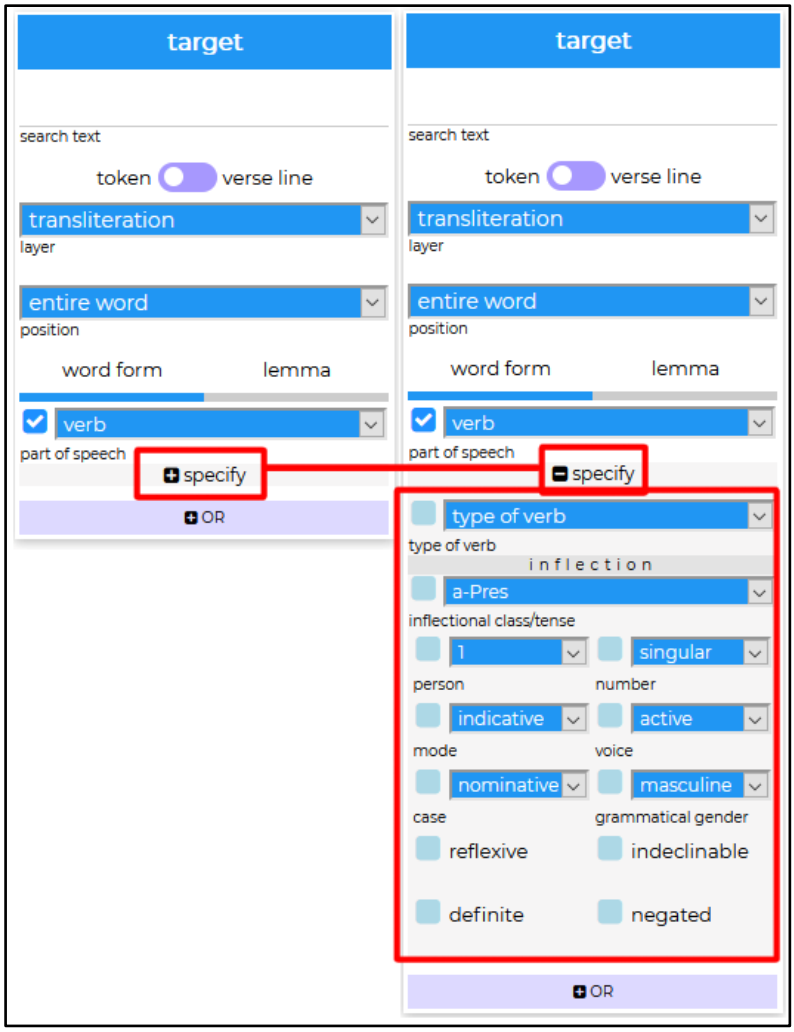
POSITION

In the **"position" dropdown** (img. 3) you can specify where the search text occurs in the target word:



Img. 3

NARROWING THE SEARCH



Img. 4

The advanced search can be narrowed to specific grammatical forms step by step. When a part of speech has been selected, clicking on SPECIFY (img. 4) will open a menu which allows the specification of a sub-class and/or morphological characteristics. The options available depend on the part of speech and on whether the search is set to "word form" or "lemma".

The tick box to the left of each option is used to filter the search:

Img. 5

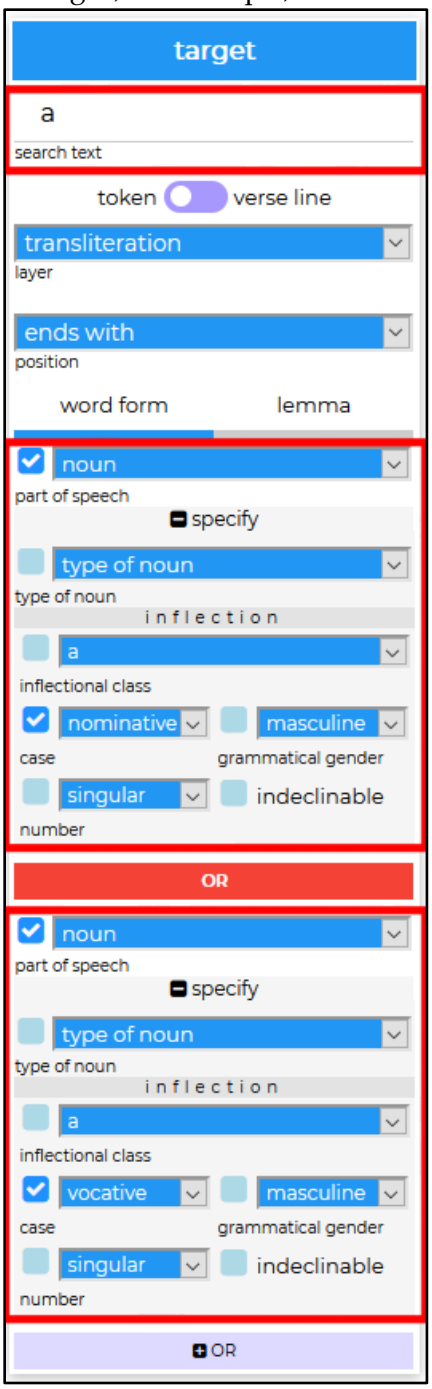
Number: any

Voice: active

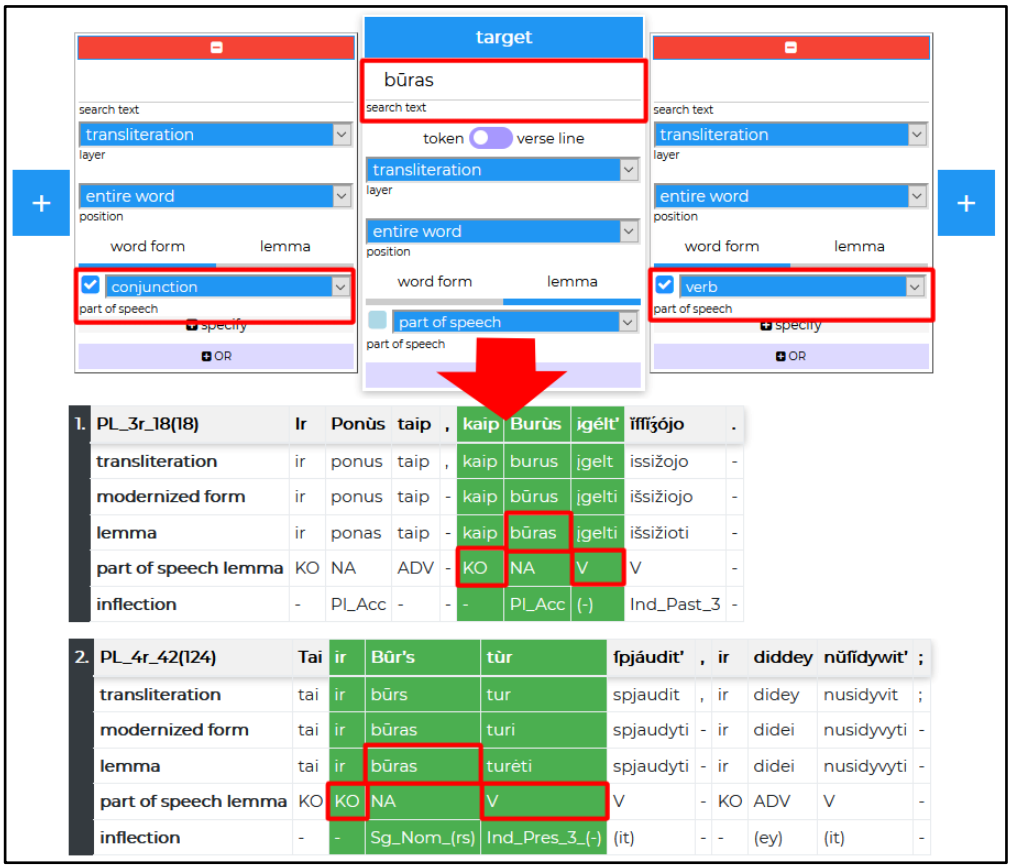
grammatical gender: **not** masculine

OR, PLUS

The OR button is used to search for multiple formal properties of the same category simultaneously. The search in img. 6, for example, will find **all nouns which end in *a* and are in the nominative OR the vocative case**.



Img. 6



Img. 7

The + buttons either side of the search box open additional search boxes. This is used to search for **word sequences** – the order of the boxes corresponds to the position in the word sequence. The search in img. 7 will thus find all sequences of **a conjunction, the lemma *būras* and a verb, in that order**.[1]

THE VERSE LINE SEARCH

In "**verse line**" mode the search checks the **entire line of verse** for the string specified and outputs all matching lines. **%** is used as a placeholder for any number of any character. A search for **%au saul%**, for example, will find the first verse of PL, „J**au saul**elė vėl atkopdama budino svietą".

[1] Note that punctuation marks are treated as tokens. For the purposes of the search, the sequence „*Sveiks , dieve duok*" consists of **four**, not three, tokens. You would thus need to search for the sequence adjective–**[blank]**–noun–verb in order to find this phrase.

Troubleshooting, special cases

In case of problems you may want to consult the troubleshooting page. Below is a list of some peculiarities of the data structure which will need to be kept in mind in order to effectively and reliably use the serach.

Verbs with alternate infinitives

For verbs in -inti/-yti the lemma entry lists two infinitives, separated by a comma without a space: "Xinti,Xyti". For example, the lemma of RG_40_5(153) *rágĭnŏ* is "raginti,ragyti". The search treats such entries as one "word" and so a search for "entire word: raginit" or "entire word: ragyti" will yield no results. To find entries of this kind, you will have to **search using "contains" rather than "entire word".**

Ambiguous forms

When a form may belong to multiple lemmata, both options are listed, separated by a slash (e.g. RG 70_4(722) *stůpu*, lemma "stuopa/stuopas"). The form's grammatical properties are listed in an analogous way – thus *stuopa/stuopas* has "o_Fem/a_Masc" as its *morphology – form* entry. **A lemma search for "entire word: *stoupa*" will not find this match**, you would need to search for "contains: *stuopa*".

Reflexives

Reflexive forms are lemmatized as reflexives. Thus, the lemma of WD_16r_42(124) *ìmmăĵĭ̃* is "imtis" (and not "imti"). To find both the reflexive and non-reflexive occurrences of a lemma you will need to run either two separate searches (for "imti" and "imtis") or search using "contains" ("contains: *imti*" will also find *imtis*).

Reflexives negated with the prefix *ne-* receive a "double" lemmatization – first with the negation, then without, separated by a slash. Thus the lemma entry for PL 10v 4(631) *nĕĵĭgėdi* is "nesigėdėtis/gėdėtis".

The prefix *be-* receives special treatment in the case of reflexives. For non-reflexive verbs it is lemmatized and glossed separately (cf. RG_40_28(176) *bĕválgant*, modernized form "be" + "valgant"). With reflexives however it is **not** separated (cf. RG 40_28(176) *bĕsĭdzaùgiant*, modernized form "besidžiaugiant", one word), because the resulting "trunk" form would otherwise be ungrammatical ((†*sidžiaugiant*).

Possessive pronouns

The uninflected possessive pronouns *mano*, *tavo* and *savo* receive a "double" lemmatization – first as the reflexive pronoun, then as the corresponding personal pronoun, separated by a slash without a space. Thus the lemma of *mano* is "mano/aš, of *tavo*: "tavo/tu", of *savo*: "savo/savęs".

nemokėti, nerimti, netekti, netikti

The above verbs have a special meaning in their negated form and are thus lemmatized *with* the negation prefix. Thus the lemma of RG_64_33(637) *nĕtìnkat* "netikti", not "tikti".

Capitalization

The search is case-sensitive. Lemmata (with the exception of proper names) and words in the transliterated and modernized text layers are always uncapitalized. But when searching the **original** text layer using the advanced search, the difference may be relevant: A search for "ar" here will find only occurrences within sentences, "Ar" only at the start of sentences.

TEXTOLOGICAL NOTES, MISPRINTS

The texts that have survived in manuscript form have been comprehensively textologically annotated in the SLIEKKAS project. These annotations have therefore not been duplicated in CorDon.
The remaining texts are based on Georg H. F. Nesselmann's print edition (1896). The occasional printing errors in these texts have been corrected on the transliteration layer. First the erroneous form is listed, then, separated by a comma, the correct form (cf. RG 36 37(109) *gìdirt*, transliteration "gidirt,girdit").