TROVA - Annotation Content Search

The latest version can be found at: http://tla.mpi.nl/tools/tla-tools/trova/

This manual was last updated in May 2013

The Language Archive, MPI for Psycholinguistics, Nijmegen, The Netherlands

TROVA - Annotation Content Search: version 1.4

The latest version can be found at: http://tla.mpi.nl/tools/tla-tools/trova/

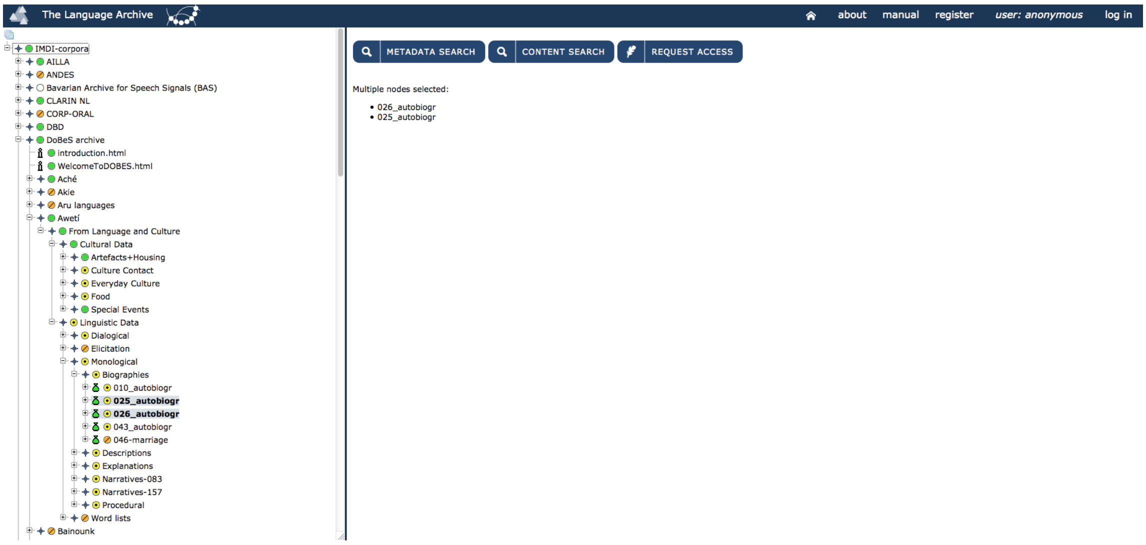
This manual was last updated in May 2013

Table of Contents

Chapter 1. TROVA: Contents search

If you want to perform a detailed search over multiple annotation files in different formats (including: EAF, SHOEBOX, TOOLBOX, CHAT, SUBRIP and TEXTGRID) you can use the **content search** function. This allows you to restrict the search to certain tiers (i.e. the different layers present in a hierarchy-structured annotation file), to use regular expressions, etc., while examining multiple annotation files at once.

To include multiple files in the search domain, select one or more nodes in the tree-view of the IMDI browser by clicking on the nodes while holding down the CTRL key on the keyboard. If you click on the button content search, you will include all annotation files under the selected corpus nodes in the search domain. Alternatively, you can select a parent node, in which case all the child nodes will be included in the search.

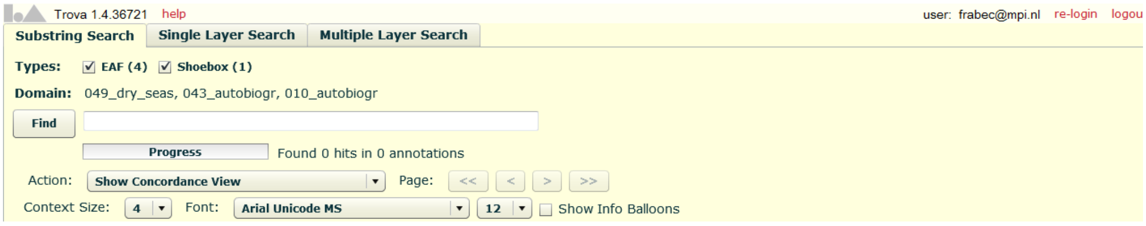


Multiple nodes selected:

- 026_autobiogr
- 025_autobiogr

Figure 1.1. Adding nodes to the search engine

A new browser frame for the search engine will appear with the search domain displayed at the top of each of the three search tabs:



Types: ☑ EAF (4) ☑ Shoebox (1)

Domain: 049_dry_seas, 043_autobiogr, 010_autobiogr

Found 0 hits in 0 annotations

Action: Show Concordance View Page:

Context Size: 4 Font: Arial Unicode MS 12 □ Show Info Balloons

Figure 1.2. Search Engine

The voice **Domain** shows the name of the selected node(s) in the archive, whereas **Types** shows all the types of annotation files in the domain of that /those node(s). You can include or exclude types of annotation files in the search by respectively checking or unchecking them. Remember that, in order to be able to see the various annotation files, you have to log in first.

It is possible that you are not allowed to read files from certain parts of the domain due to AMS rules (http://tla.mpi.nl/tools/tla-tools/ams/). In that case all searchable files from those parts are excluded from the annotation content search. As a result, the search domain contains fewer files.

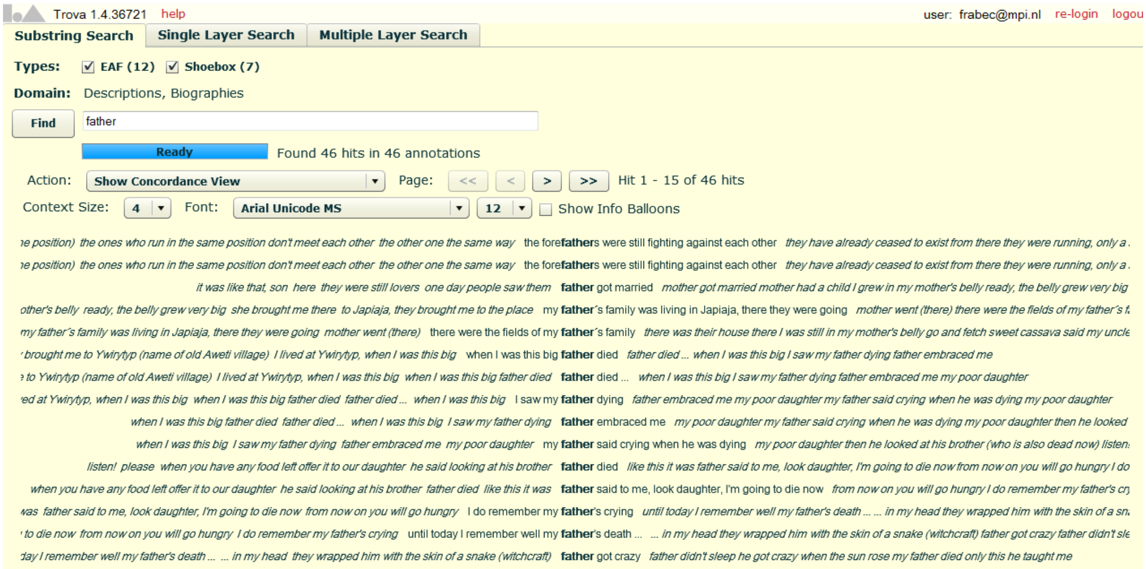
Please remember that even if you have received access to the resources through AMS, it could take up to one day for the changes to be recognised by TROVA as well.

As can be seen from figure 1.2 above, there are three tabs offering different kinds of search:

- Substring Search: it finds all annotations in which the search string occurs.
- Single Layer Search: it finds all annotations or N-grams (i.e. search strings of more than one word, either consecutive or not) in which either the search string or the regular expression occurs, both case sensitive and insensitive and possibly restricted to one (type of) tier.
- Multiple Layer Search: it finds annotations within one single tier but also in three related tiers at the same time. You can use multiple search strings or regular expressions and make constraints on duration and time slot as well as constraints on how the search strings are to be combined.

1.1. Substring Search Tab

This tab offers the simplest search. It just asks for a search string. After entering the search string you can click on Find (or press **Enter**) to start the search process. This will result in a page like the one below:



user: frabec@mpi.nl

Types: ☑ EAF (12) ☑ Shoebox (7)

Domain: Descriptions, Biographies

father

Ready Found 46 hits in 46 annotations

Action: Show Concordance View

Hit 1 - 15 of 46 hits

Context Size: 4 Font: Arial Unicode MS 12 □ Show Info Balloons

e position) the ones who run in the same position don't meet each other the other one the same way the fore**father**s were still fighting against each other *they have already ceased to exist from there they were running, only a .*

e position) the ones who run in the same position don't meet each other the other one the same way the fore**father**s were still fighting against each other *they have already ceased to exist from there they were running, only a .*

it was like that, son here they were still lovers one day people saw them **father** got married *mother got married mother had a child I grew in my mother's belly ready, the belly grew very big*

other's belly ready, the belly grew very big she brought me there to Japiaja, they brought me to the place my **father**´s family was living in Japiaja, there they were going *mother went (there) there were the fields of my father´s f.*

my father´s family was living in Japiaja, there they were going mother went (there) there were the fields of my **father**´s family *there was their house there I was still in my mother's belly go and fetch sweet cassava said my uncle*

brought me to Ywirytyp (name of old Aweti village) I lived at Ywirytyp, when I was this big when I was this big **father** died *father died ... when I was this big I saw my father dying father embraced me*

e to Ywirytyp (name of old Aweti village) I lived at Ywirytyp, when I was this big when I was this big father died **father** died ... *when I was this big I saw my father dying father embraced me my poor daughter*

ed at Ywirytyp, when I was this big when I was this big father died father died ... when I was this big I saw my **father** dying *father embraced me my poor daughter my father said crying when he was dying my poor daughter*

when I was this big father died father died ... when I was this big I saw my father dying **father** embraced me *my poor daughter my father said crying when he was dying my poor daughter then he looked*

when I was this big I saw my father dying father embraced me my poor daughter my **father** said crying when he was dying *my poor daughter then he looked at his brother (who is also dead now) listen.*

listen! please when you have any food left offer it to our daughter he said looking at his brother **father** died *like this it was father said to me, look daughter, I'm going to die now from now on you will go hungry I do*

when you have any food left offer it to our daughter he said looking at his brother father died like this it was **father** said to me, look daughter, I'm going to die now *from now on you will go hungry I do remember my father's cry*

was father said to me, look daughter, I'm going to die now from now on you will go hungry I do remember my **father**'s crying *until today I remember well my father's death ... ... in my head they wrapped him with the skin of a sn.*

to die now from now on you will go hungry I do remember my father's crying until today I remember well my **father**'s death ... *... in my head they wrapped him with the skin of a snake (witchcraft) father got crazy father didn't sle*

day I remember well my father's death ... ... in my head they wrapped him with the skin of a snake (witchcraft) **father** got crazy *father didn't sleep he got crazy when the sun rose my father died only this he taught me*

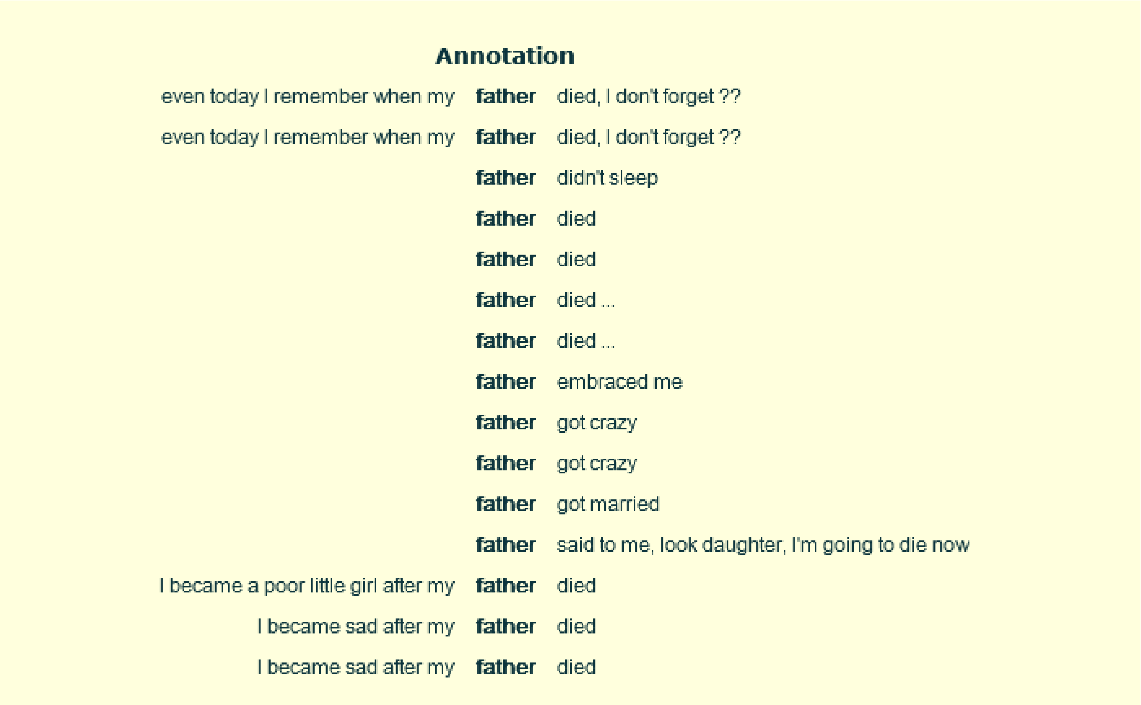
Figure 1.3. Substring Search Results

The result page above shows the annotations containing the search string plus some annotations in the context printed in italic. The default number of annotations in the context is four on both sides, but you can change it through the **Context size** menu. When the number of hits exceeds the maximum number the window can contain, you can view the rest of the hits by clicking the **<** and **>** buttons that appear above the list of hits. Besides the ones just described, there are other options you can choose from.

From the **Action** menu you can select:

1) Show Concordance View: it is the default mode in which results are shown. The search strings is in bold and has some context on both sides.

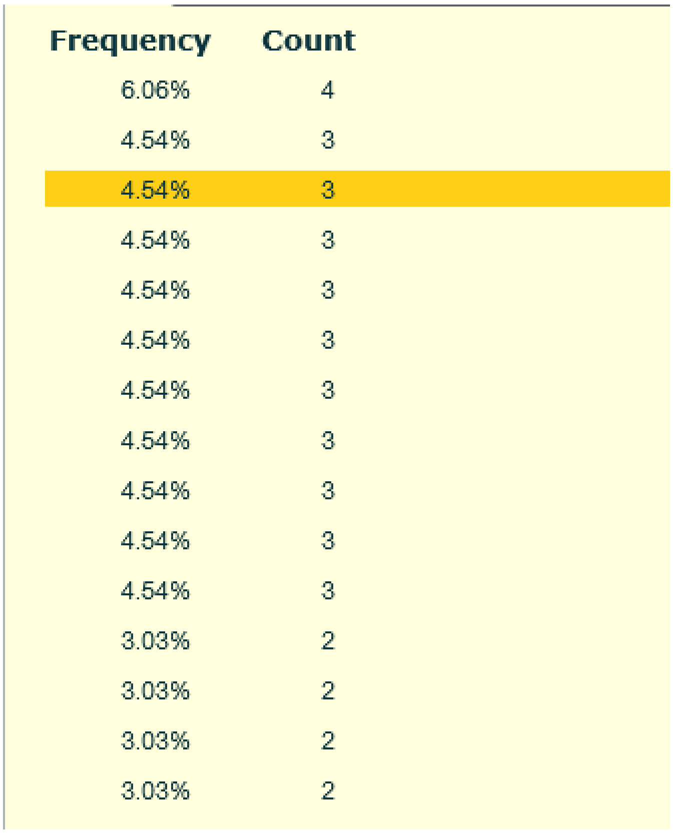
2) Show Frequency View Sorted by Annotation: the results are shown in the right bottom half of the page in alphabetical order (see the first word of the annotation).



	Annotation	
even today I remember when my	**father**	died, I don't forget ??
even today I remember when my	**father**	died, I don't forget ??
	father	didn't sleep
	father	died
	father	died
	father	died ...
	father	died ...
	father	embraced me
	father	got crazy
	father	got crazy
	father	got married
	father	said to me, look daughter, I'm going to die now
I became a poor little girl after my	**father**	died
I became sad after my	**father**	died
I became sad after my	**father**	died

Figure 1.4. Frequency View Sorted by Annotation

3) Show Frequency View Sorted by Frequency: the results are shown in the left, bottom half of the page, under percentage form, in a decreasing order.



Frequency	Count
6.06%	4
4.54%	3
4.54%	3
4.54%	3
4.54%	3
4.54%	3
4.54%	3
4.54%	3
4.54%	3
4.54%	3
4.54%	3
3.03%	2
3.03%	2
3.03%	2
3.03%	2

Figure 1.5. Frequency View Sorted by Frequency

4) Show Selected Hit in Transcription: as the name says, first of all select one of the results, and then select the option itself (otherwise simply double click the result you need). You will be redirected to Annex interface, where you will see the annotation in the Timeline viewer [see section 2.2.1. of Annex manual; use the following link: http://www.mpi.nl/corpus/html/annex/ch02s02s01.html].

5) Show Selected Hit in Corpus Tree: like with option 4), first select the result you need, and then this option. You will be redirected to the IMDI browser interface, where you will see: a) the resource to which the annotation belongs highlighted in grey in the IMDI tree; b) some information about such resource in the main content panel.

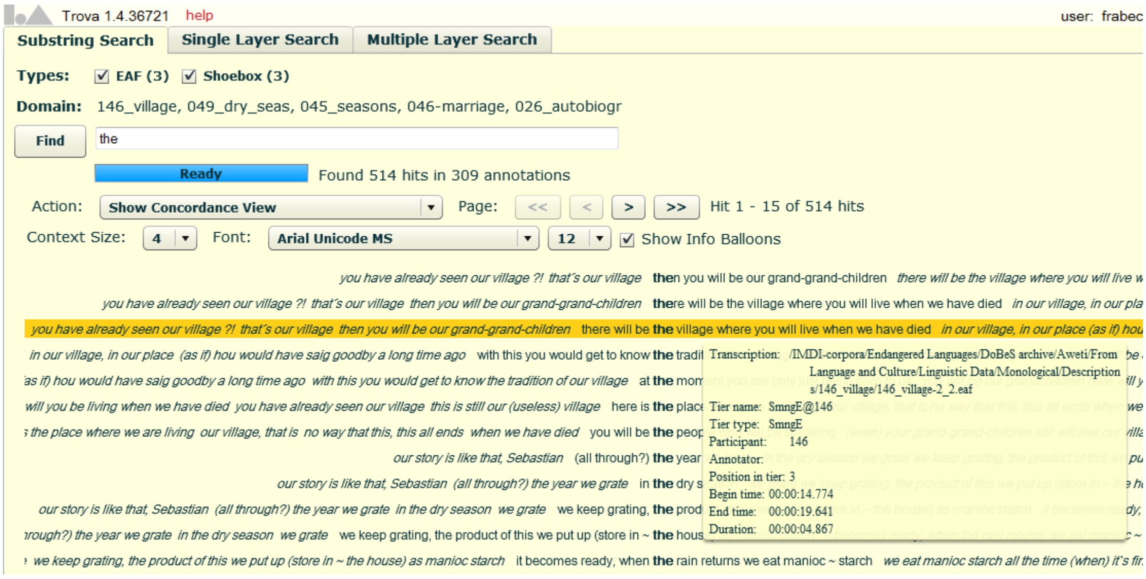
6) **Save Hits**: this option allows you to save the results you have obtained from your query. Once selected, it will open a dialogue window asking you to choose: a) the export format (either UTF-8 or UTF-16); b) the fields to export (e.g. begin and end time, tier type and name, etc.).

Note

If you are in doubt about which one to choose between UTF-8 and UTF-16 because you are not very familiar with character encoding, opt for UTF-8.

Moreover, below the **Action** drop-down menu, you can find a **Font** menu, which allows you to choose the style of writing and the size in which the results will be shown.

Finally, the option **Show Info Balloons**, which is displayed on the right of the **Font** menu, allows you to see some additional information about the annotations you have searched for. If you check this option, and then place the mouse cursor on one of the results, you will see a yellow balloon appearing, containing some information about that result, like for example the transcription link, the tier name and type, the begin and end time, etc. See figure 1.4. below.



Types: ☑ EAF (3) ☑ Shoebox (3)

Domain: 146_village, 049_dry_seas, 045_seasons, 046-marriage, 026_autobiogr

the

Found 514 hits in 309 annotations

Action: Show Concordance View ▾ Page: Hit 1 - 15 of 514 hits

Context Size: 4 ▾ Font: Arial Unicode MS ▾ 12 ▾ ☑ Show Info Balloons

you have already seen our village ?! that´s our village **the**n you will be our grand-grand-children *there will be the village where you will live w*

you have already seen our village ?! that´s our village then you will be our grand-grand-children **the**re will be the village where you will live when we have died *in our village, in our pla*

you have already seen our village ?! that´s our village then you will be our grand-grand-children there will be **the** village where you will live when we have died *in our village, in our place (as if) hou*

in our village, in our place (as if) hou would have saig goodby a long time ago with this you would get to know **the** tradi

as if) hou would have saig goodby a long time ago with this you would get to know the tradition of our village at **the** mo

will you be living when we have died you have already seen our village this is still our (useless) village here is **the** place

the place where we are living our village, that is no way that this, this all ends when we have died you will be **the** peop

our story is like that, Sebastian (all through?) **the** year

our story is like that, Sebastian (all through?) the year we grate in **the** dry s

our story is like that, Sebastian (all through?) the year we grate in the dry season we grate we keep grating, **the** prod

rough?) the year we grate in the dry season we grate we keep grating, the product of this we put up (store in ~ **the** hous

we keep grating, the product of this we put up (store in ~ the house) as manioc starch it becomes ready, when **the** rain returns we eat manioc ~ starch *we eat manioc starch all the time (when) it´s fir*

Transcription: /IMDI-corpora/Endangered Languages/DoBeS archive/Aweti/From Language and Culture/Linguistic Data/Monological/Description s/146_village/146_village-2_2.eaf
Tier name: SmngE@146
Tier type: SmngE
Participant: 146
Annotator:
Position in tier: 3
Begin time: 00:00:14.774
End time: 00:00:19.641
Duration: 00:00:04.867

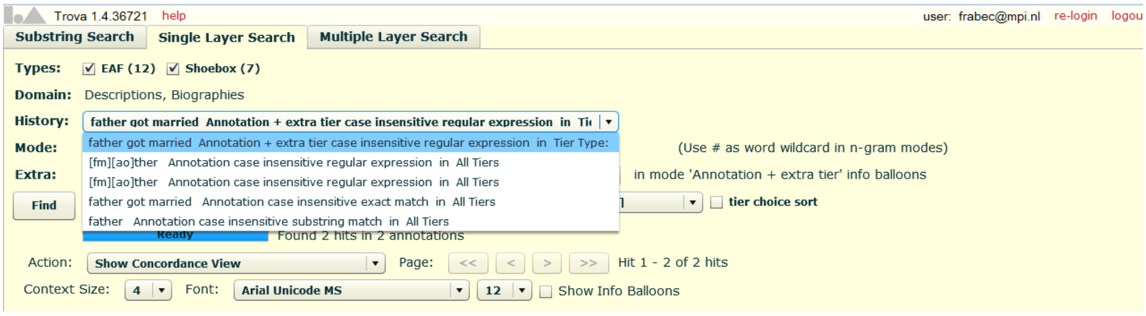
Figure 1.6. Show Info Balloons

Note

All the features seen above, related to the **Substring Search**, are also present (with the same functions) in the **Single Layer Search** and in the **Multiple Layer Search**.

1.2. Single Layer Search Tab

The Single Layer Search tab offers a more elaborate search than the Substring Search tab. The first thing that is different from the Substring Search tab is that the Single Layer Search tab has a query history. Clicking on the **History** roll-out menu, you can have a look at the previous queries made during that ongoing session. See the figure below.



Types: ☑ EAF (12) ☑ Shoebox (7)

Domain: Descriptions, Biographies

History: father got married Annotation + extra tier case insensitive regular expression in Ti

father got married Annotation + extra tier case insensitive regular expression in Tier Type:
[fm][ao]ther Annotation case insensitive regular expression in All Tiers
[fm][ao]ther Annotation case insensitive regular expression in All Tiers
father got married Annotation case insensitive exact match in All Tiers
father Annotation case insensitive substring match in All Tiers

Mode: (Use # as word wildcard in n-gram modes)

Extra: in mode 'Annotation + extra tier' info balloons

Found 2 hits in 2 annotations

Action: Show Concordance View

Hit 1 - 2 of 2 hits

Context Size: 4 Font: Arial Unicode MS 12 □ Show Info Balloons

Figure 1.7. Query History in Single Layer Search

Furthermore, this search tab offers different modes to restrict the search. The first mode allows you to choose the way in which the results are shown. There are four options:

- Annotation: the search string is part of or exactly matches an annotation.
- N-gram over annotations: each element of the search string (the elements must be divided by spaces) is part of or exactly matches each one of several consecutive annotations.
- N-gram within annotation: the various elements of the search string (the elements must be divided by spaces) are part of or exactly match one single annotation.
- **Annotation + extra tier**: with this option you can search for annotations in two tiers at the same time. Once selected all the necessary constraints, the first tier will be shown in the results, whereas the additional tier will appear into the information balloon when hovering over such results. When you choose this option, another menu will become available, including the following options:
  - **Fully aligned, any tier**: the begin and end time of both annotations are the same. In this case you do not specify the type of the additional tier.
  - **Overlapping, any tier**: part of both annotations overlap on the time axis. Here again you do not specify the type of the additional tier.
  - **Fully aligned, sibling tier**: the begin and end time of both annotations are the same, but the two tiers in which the annotations are contained are siblings, i.e. they belong to the same parent tier.
  - **Overlapping, sibling tier**: part of both annotations overlap on the time axis. The two tiers containing the searched annotations are siblings, i.e. they belong to the same parent tier.
  - **Fully aligned, child tier**: the two annotations have the same time span, but the additional one, contained in the extra tier, is child of the other one.
  - **Overlapping, child tier**: part of both annotations overlap on the time axis, but the additional one, contained in the extra tier, is child of the other one.

The second mode offers the straightforward distinction between case sensitive and case insensitive search.

The third mode allows you to choose whether the annotation found should contain the search string (substring match), whether the annotation should exactly match the search string (exact match) or whether some regular expression should be used in the match (regular expression) [for further information about regular expressions see section 1.4 below]. This mode also contains another option, which is called **variable match**, but since it has little use here in the **Single Layer Search**, it will be thoroughly explained later on, when it comes to the **Multiple Layer Search**.

Finally, you can restrict the search to one tier in particular, choosing among tier types, tier names, participants or annotators (if present), all of which can appear in a decreasing order, from the most to the least used, if you check the option **tier choice sort**.

Wild cards and negation

When you choose an N-gram to be the form of the result, you can use two more options: a wild card and/or a negation. The wild card takes the form of a # sign. For instance, the search string **`the # man`** with the mode N-gram over annotations would return three annotations per hit: the first annotation contains `the` (or exactly matches it, if the mode exact match is chosen), the second annotation may contain anything due to the use of the wild card, and the third annotation contains, or exactly matches, `man`. If the mode N-gram within annotation is chosen, each hit contains one annotation. This one, in turn, contains the search string with all the possible combinations due to the # sign.

If you want to find N-grams where a hit matches anything but one string in particular, you can use the negation operator NOT(...), entering in brackets the search string not to be matched. For instance, the search string **`the NOT(strange) man`** would return 3-grams in the same way as described above, but the hits where the second annotation matches, or contains, *strange* are left out.

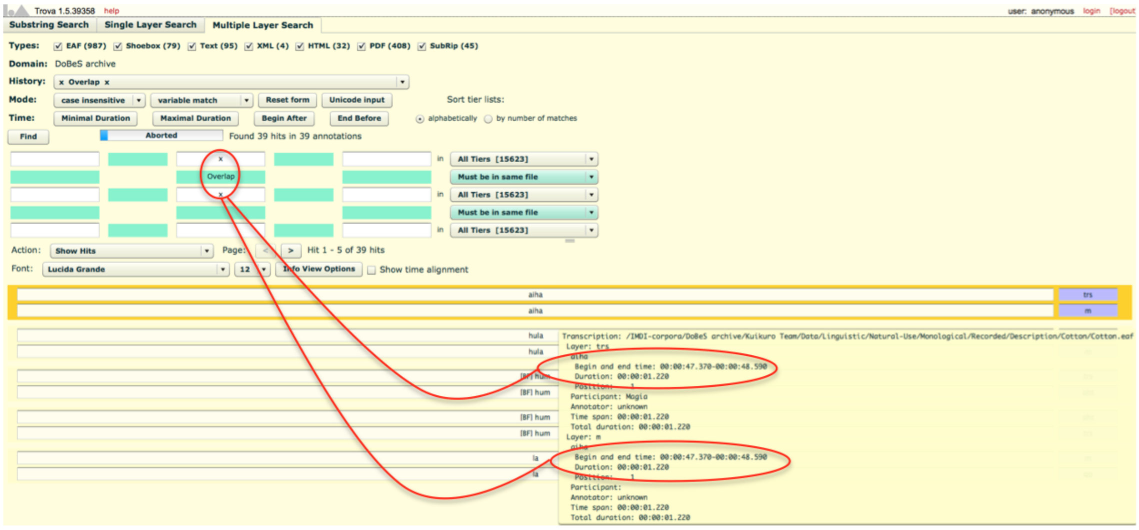
1.3. Multiple Layer Search Tab

The Multiple Layer Search tab houses the most comprehensive search in Trova.

Similarly to the Single Layer Search tab, it has a query History menu, which allows you to go back to the previous queries you have made during that ongoing session. Also the mode case sensitive/case insensitive is the same as the one contained in the Single Layer Search tab.

The first new element is the Reset form button. Clicking this button will clear all the strings entered in order to make a query.

Besides, a new option has been included into the menu containing all the different types of matches (i.e. substring match, exact match, regular expression): **variable match**. As the name says, it has to do with mathematical variables, and it can be used every time you want to search for two or more annotations, contained in two or more different tiers, reporting the same text and/or the same time alignment. See the example in the figure below.



Types: ☑ EAF (987) ☑ Shoebox (79) ☑ Text (95) ☑ XML (4) ☑ HTML (32) ☑ PDF (408) ☑ SubRip (45)

Domain: DoBeS archive

History: x Overlap x

Mode: case insensitive | variable match

Sort tier lists: ● alphabetically ○ by number of matches

Found 39 hits in 39 annotations

x in All Tiers [15623]
Overlap — Must be in same file
x in All Tiers [15623]
Must be in same file
in All Tiers [15623]

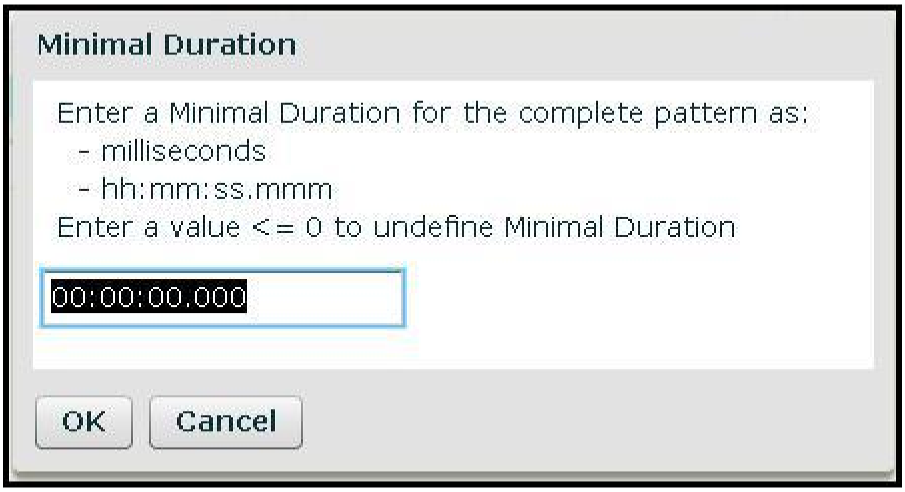
Action: Show Hits Page: < > Hit 1 - 5 of 39 hits

Font: Lucida Grande 12 □ Show time alignment

```
Transcription: /IMDI-corpora/DoBeS archive/Kuikuro Team/Data/Linguistic/Natural-Use/Monological/Recorded/Description/Cotton/Cotton.eaf
 Layer: trs
  aiha
   Begin and end time: 00:00:47.370-00:00:48.590
   Duration: 00:00:01.220
   Position:    1
  Participant: Magia
  Annotator: unknown
  Time span: 00:00:01.220
  Total duration: 00:00:01.220
 Layer: m
  aiha
   Begin and end time: 00:00:47.370-00:00:48.590
   Duration: 00:00:01.220
   Position:    1
  Participant:
  Annotator: unknown
  Time span: 00:00:01.220
  Total duration: 00:00:01.220
```

Figure 1.8. Example Variable Match

The buttons Minimal Duration and Maximal Duration enables you to constrict the minimal and maximal temporal duration of each result (please note in the window below that it says "Enter a Minimal/Maximal Duration for the complete pattern", which means that if you have filled in all the three strings, you will have to enter the whole duration, and not the duration of one single annotation). When you click on one of the buttons, a dialogue window appears:



Minimal Duration

Enter a Minimal Duration for the complete pattern as:
- milliseconds
- hh:mm:ss.mmm

Enter a value <= 0 to undefine Minimal Duration

00:00:00.000

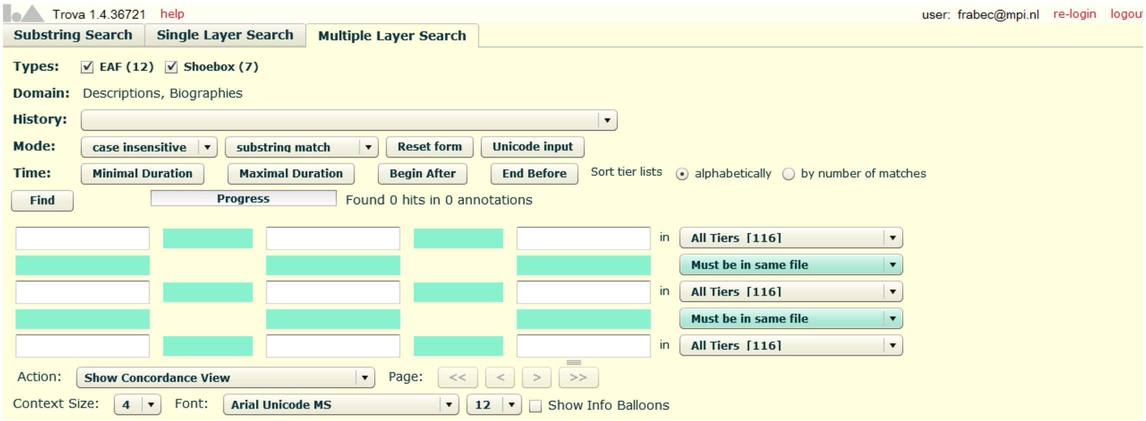
Figure 1.9. Minimal Duration

Here you can enter the minimal and/or maximal duration as the total number of milliseconds or in the form of hours:minutes:seconds.milliseconds. A value of 0 milliseconds or 00:00:00.000 yields as undefined. Searching for annotations with a maximum duration being less then the minimum duration is impossible. Hence, entering conflicting values results in an error message saying that the combination is impossible. After entering a correct duration, this will be displayed in the corresponding button.

The buttons Begin After and End Before give a dialogue window similar to the one seen above. They allow you to restrict the annotations in the results so that they begin after a certain time and end before a certain time. Entering a Begin After time that is higher than the End Before time results in an error message saying that the query made is impossible. After entering a correct time, this will be displayed in the corresponding button.

Search string and constraints

Beneath all the buttons and functions discussed above, you will find a table consisting of white and green fields. Search strings are entered in the white fields, while a green field between two non-empty white fields must contain a constraint.



Trova 1.4.36721 help

user: frabec@mpi.nl re-login logou

Types: ☑ EAF (12) ☑ Shoebox (7)

Domain: Descriptions, Biographies

History:

Mode: case insensitive, substring match

Time: Sort tier lists ● alphabetically ○ by number of matches

Found 0 hits in 0 annotations

in All Tiers [116]

Must be in same file

in All Tiers [116]

Must be in same file

in All Tiers [116]

Action: Show Concordance View

Context Size: 4 Font: Arial Unicode MS 12 □ Show Info Balloons

Figure 1.10. Multiple Layer Search

The fields on one row have to do with the search strings and constraints that have to be matched by annotations contained in one tier. The reason for having three rows in the query table is that the search engine may find annotations contained in three tiers as one hit. Furthermore, it is possible to restrict the search to one type of tier for each row by choosing the appropriate option in the pull-down menu on the right of each row.

Let us first take a look at search strings and constraints in one row. If you enter two search strings in two white fields separated by a green field, you must fill in that green field, i.e. choose a constraint. Double clicking on the green field will open a context menu offering the following constraints:

- = N annotations: between the annotations containing the two search strings, there must be exactly N annotations.
- > N annotations: between the annotations containing the two search strings, there must be more than N annotations.
- < N annotations: between the annotations containing the two search strings, there must be less than N annotations.
- = X milliseconds: between the annotations containing the two search strings, there must be exactly X milliseconds.
- > X milliseconds: between the annotations containing the two search strings, there must be more than X milliseconds.
- < X milliseconds: between the annotations containing the two search strings, there must be less than X milliseconds.
- No constraints: there are no constraints between the annotations.
- Clear: it clears the constraint previously chosen.

When you click on Find and there is an empty constraint between two non-empty search string fields, you will get an error message. You will also get an error message if there is an empty search string field and empty constraint fields between two non-empty search string fields.

As we saw earlier, the search mechanism on this tab can construct a query within up to three tiers. Besides the constraints on the annotations present in one tier, you can also apply constraints on annotations contained in different tiers. This means that, taking into consideration the constraints entered in the green fields, the search engine will look for annotations which are contained in different tiers, and which match the search strings entered in two (or three) different rows of the query table.

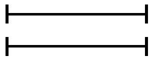
As you can see in the figure above, besides choosing the type of tier, you can also choose among the following options, which are contained in the green drop down menus in between the tier types menus:

- **Must be in same file**: by choosing this option Trova will search for two tiers belonging to the same annotation file, but without a specified relationship between them.
- **Must be parent and child**: by choosing this option Trova will search for two tiers, one being child of the other. Please note that if you look at the annotations in the Timeline view mode in Annex, the parent tier is generally located in the lower positions, whereas when it comes to make a query, it has to go in the top search strings.
- **Must have same parent**: by choosing this option Trova will search for two sibling tiers, which means they belong to the same parent tier.
- **Must have same participant**: by choosing this option Trova will search for two (or more) tiers whose content has been orally told by the same person. The identity (and the identifying number) of the participants can be seen on the IMDI Browser page once you have selected the required node(s).

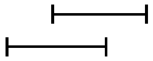
  Multiple constraints can be selected from this menu. All you have to do is holding down the Ctrl key on the keyboard while selecting the needed items.

Double clicking the green field between two search strings opens a context menu with the following constraints:

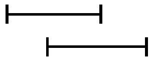
- Fully aligned: the begin time and end time of both annotations are the same;



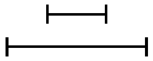
- Overlap: part of both annotations overlap;
- Left overlap: the begin time and end time of the annotation matching the lower search string lie *before* the begin time and end time of the annotation matching the upper search string;



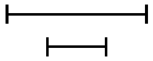
- Right overlap: the begin time and end time of the annotation matching the lower search string lie *after* the begin time and end time of the annotation matching the upper search string;



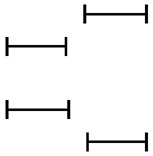
- Surrounding: the begin time of the annotation matching the lower search string lies before the begin time of the annotation matching the upper search string; and the end time of the annotation matching the lower search string lies after the end time of the annotation matching the upper search string;



- Within: the begin time of the annotation matching the lower search string lies after the begin time of the annotation matching the upper search string; and end time of the annotation matching the lower search string lies before the end time of the annotation matching the upper search string;



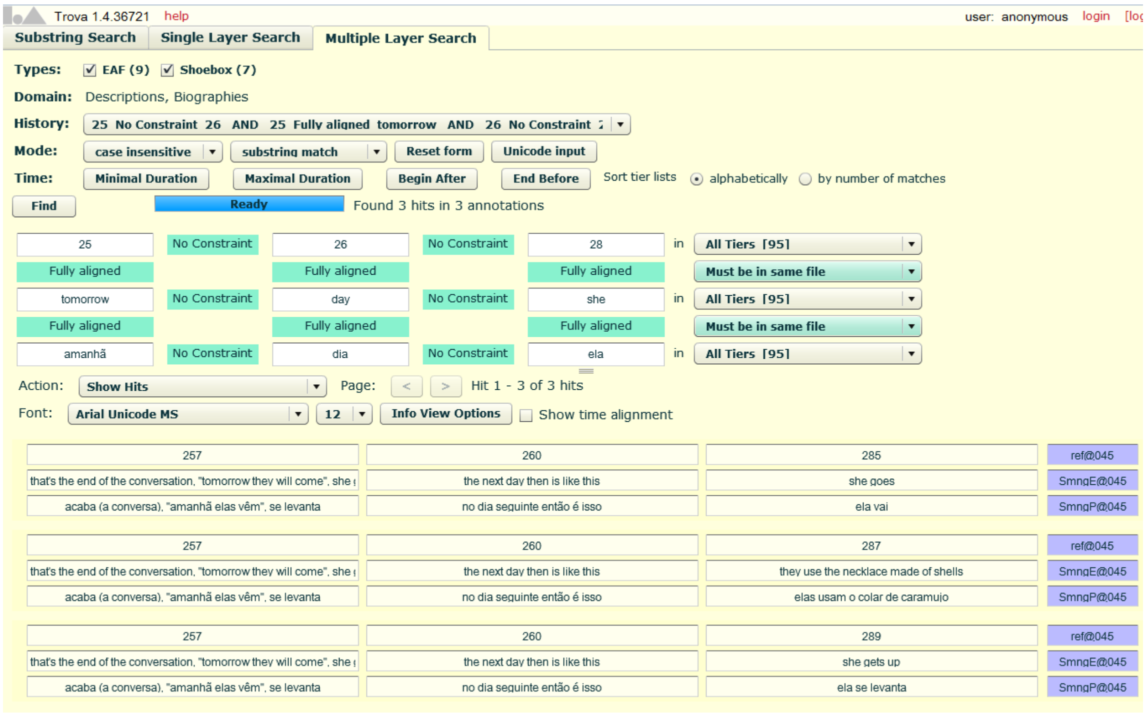
- No overlap: the begin time of the annotation matching one of the search strings lies after the end time of the annotation matching the other search string;



- begin time - begin time = X milliseconds: the begin time of the annotation matching the upper search string must lie exactly X milliseconds before the begin time of the annotation matching the lower search string;
- begin time - begin time < X milliseconds: the begin time of the annotation matching the upper search string must lie less than X milliseconds before the begin time of the annotation matching the lower search string;
- begin time - begin time > X milliseconds: the begin time of the annotation matching the upper search string must lie more than X milliseconds before the begin time of the annotation matching the lower search string;
- begin time - end time = X milliseconds: the begin time of the annotation matching the upper search string must lie exactly X milliseconds before the end time of the annotation matching the lower search string;

- begin time - end time < X milliseconds: the begin time of the annotation matching the upper search string must lie less than X milliseconds before the end time of the annotation matching the lower search string;
- begin time - end time > X milliseconds: the begin time of the annotation matching the upper search string must lie more than X milliseconds before the end time of the annotation matching the lower search string;
- end time - begin time = X milliseconds: the end time of the annotation matching the upper search string must lie exactly X milliseconds before the begin time of the annotation matching the lower search string;
- end time - begin time < X milliseconds: the end time of the annotation matching the upper search string must lie less than X milliseconds before the begin time of the annotation matching the lower search string;
- end time - begin time > X milliseconds: the end time of the annotation matching the upper search string must lie more than X milliseconds before the begin time of the annotation matching the lower search string;
- end time - end time = X milliseconds: the end time of the annotation matching the upper search string must lie exactly X milliseconds before the end time of the annotation matching the lower search string;
- end time - end time < X milliseconds: the end time of the annotation matching the upper search string must lie less than X milliseconds before the end time of the annotation matching the lower search string;
- end time - end time > X milliseconds: the end time of the annotation matching the upper search string must lie more than X milliseconds before the end time of the annotation matching the lower search string;
- No constraint: there are no constraints between the annotations or the tiers.
- Clear: it clears the constraint previously chosen.

Because the search mechanism offers the possibility to search for patterns in three tiers and there are possibly three search strings per tier, the search results also consist of nine elements per hit. The results are presented in the form of a table. In the figure below you can see an example of a query and its results. The upper row is the parent tier, the middle row is the English gloss tier, and the lower row is the Portuguese gloss tier.



Types: ☑ EAF (9) ☑ Shoebox (7)

Domain: Descriptions, Biographies

History: 25 No Constraint 26 AND 25 Fully aligned tomorrow AND 26 No Constraint 2

Mode: case insensitive ▾ substring match ▾

Sort tier lists ● alphabetically ○ by number of matches

Ready Found 3 hits in 3 annotations

25	No Constraint	26	No Constraint	28	in	All Tiers [95]
Fully aligned		Fully aligned		Fully aligned		Must be in same file
tomorrow	No Constraint	day	No Constraint	she	in	All Tiers [95]
Fully aligned		Fully aligned		Fully aligned		Must be in same file
amanhã	No Constraint	dia	No Constraint	ela	in	All Tiers [95]

Action: Show Hits ▾ Page: < > Hit 1 - 3 of 3 hits

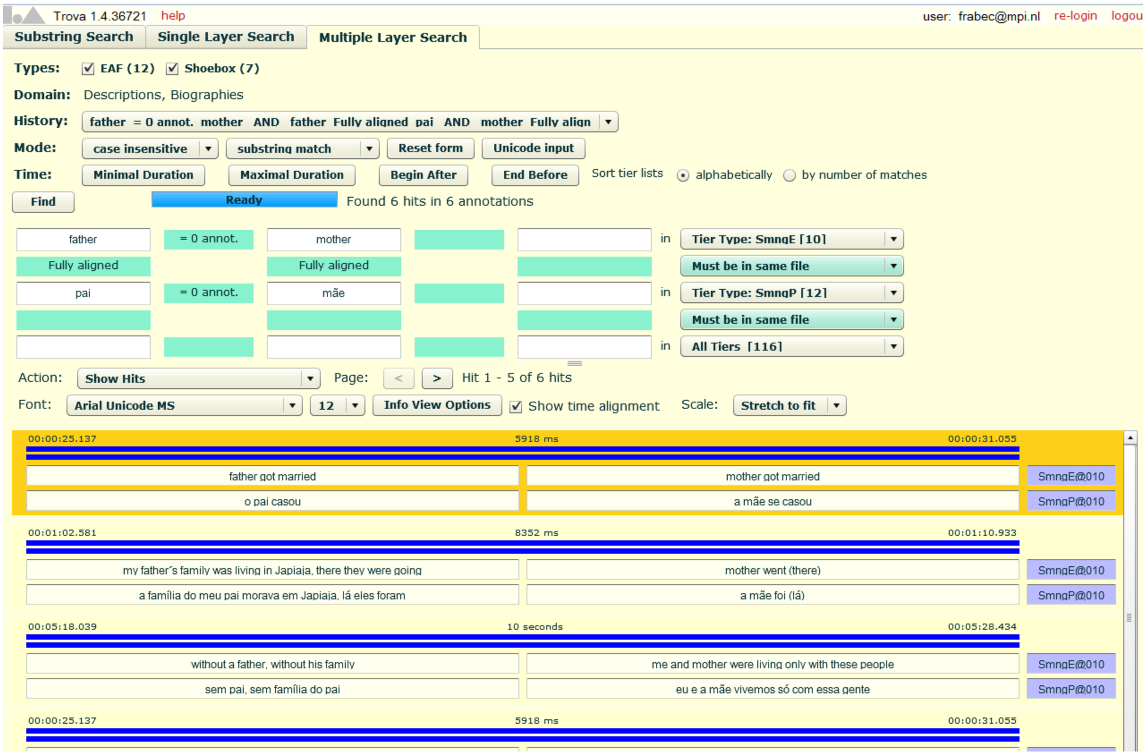
Font: Arial Unicode MS ▾ 12 ▾ □ Show time alignment

257	260	285	ref@045
that's the end of the conversation, "tomorrow they will come", she g	the next day then is like this	she goes	SmngE@045
acaba (a conversa), "amanhã elas vêm", se levanta	no dia seguinte então é isso	ela vai	SmngP@045
257	260	287	ref@045
that's the end of the conversation, "tomorrow they will come", she g	the next day then is like this	they use the necklace made of shells	SmngE@045
acaba (a conversa), "amanhã elas vêm", se levanta	no dia seguinte então é isso	elas usam o colar de caramujo	SmngP@045
257	260	289	ref@045
that's the end of the conversation, "tomorrow they will come", she g	the next day then is like this	she gets up	SmngE@045
acaba (a conversa), "amanhã elas vêm", se levanta	no dia seguinte então é isso	ela se levanta	SmngP@045

Figure 1.11. Results of Multiple Layer Search

Here in the **Multiple Layer Search** you can sort out the tier lists in two ways: 1) **alphabetically**; 2) **by number of matches** (the matches are presented in a decreasing order). In both cases, the tiers will be ordered as follows: tier type, tier name, participant (and annotator, if present).

In addition, on the right of the **Font** menu there is an option called **Show time alignment**, which appears ONLY when the search has been made over TWO (or more) different tiers. Once you have checked this option, you will see, above each one of the results, two blue bars (graphically representing the time span of the annotations searched) plus the begin and end time of the annotations, and their duration in milliseconds. See figure below.



Trova 1.4.36721 help user: frabec@mpi.nl re-login logou

Substring Search | Single Layer Search | Multiple Layer Search

Types: ☑ EAF (12) ☑ Shoebox (7)

Domain: Descriptions, Biographies

History: father = 0 annot. mother AND father Fully aligned pai AND mother Fully align

Mode: case insensitive | substring match | Reset form | Unicode input

Time: Minimal Duration | Maximal Duration | Begin After | End Before | Sort tier lists ● alphabetically ○ by number of matches

Find | Ready | Found 6 hits in 6 annotations

father	= 0 annot.	mother			in	Tier Type: SmngE [10]
Fully aligned		Fully aligned				Must be in same file
pai	= 0 annot.	mãe			in	Tier Type: SmngP [12]
						Must be in same file
					in	All Tiers [116]

Action: Show Hits | Page: < > Hit 1 - 5 of 6 hits

Font: Arial Unicode MS | 12 | Info View Options | ☑ Show time alignment | Scale: Stretch to fit

00:00:25.137 — 5918 ms — 00:00:31.055

father got married	mother got married	SmngE@010
o pai casou	a mãe se casou	SmngP@010

00:01:02.581 — 8352 ms — 00:01:10.933

my father's family was living in Japiaia, there they were going	mother went (there)	SmngE@010
a família do meu pai morava em Japiaia, lá eles foram	a mãe foi (lá)	SmngP@010

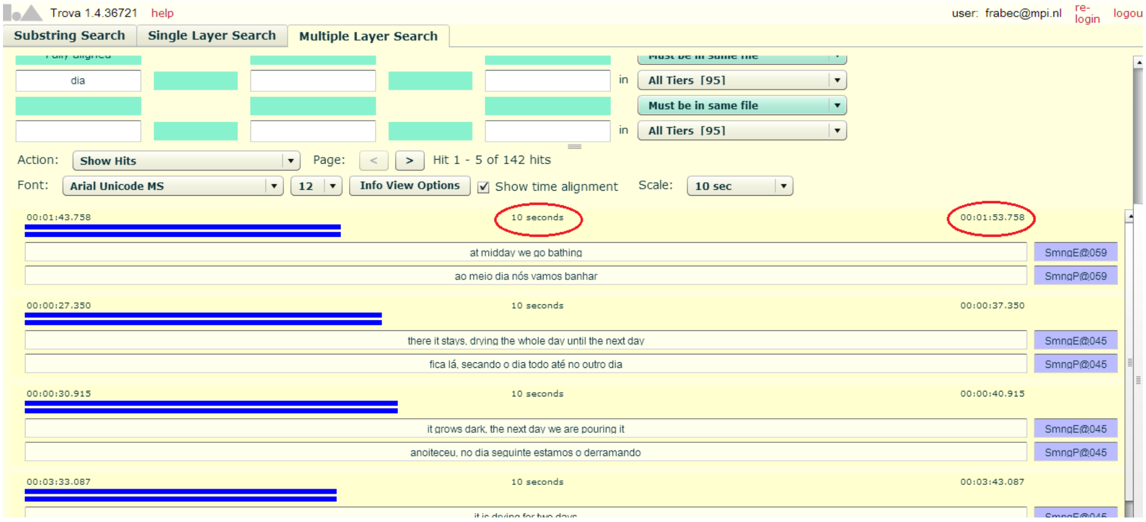
00:05:18.039 — 10 seconds — 00:05:28.434

without a father, without his family	me and mother were living only with these people	SmngE@010
sem pai, sem família do pai	eu e a mãe vivemos só com essa gente	SmngP@010

00:00:25.137 — 5918 ms — 00:00:31.055

Figure 1.12. Show Time Alignment

As soon as this option is checked, another drop down menu will open up, named **Scale**, with the item **Stretch to fit** as default option. The latter means, on the one hand, that the two blue bars will stretch as much as the page allows them to, and on the other hand that each of the result will show its own begin and end time, and its own duration time in milliseconds. If you change the scale to, for example, 10 seconds, you will see that the two bars will have changed their length accordingly (thus making the visualisation easier and quicker), that all the results will report a complete duration of ten seconds, and that the end time of the annotations will correspond to the begin time plus ten seconds. See figure below.



dia

Hit 1 - 5 of 142 hits

☑ Show time alignment

00:01:43.758	10 seconds	00:01:53.758

at midday we go bathing — SmngE@059

ao meio dia nós vamos banhar — SmngP@059

00:00:27.350	10 seconds	00:00:37.350

there it stays, drying the whole day until the next day — SmngE@045

fica lá, secando o dia todo até no outro dia — SmngP@045

00:00:30.915	10 seconds	00:00:40.915

it grows dark, the next day we are pouring it — SmngE@045

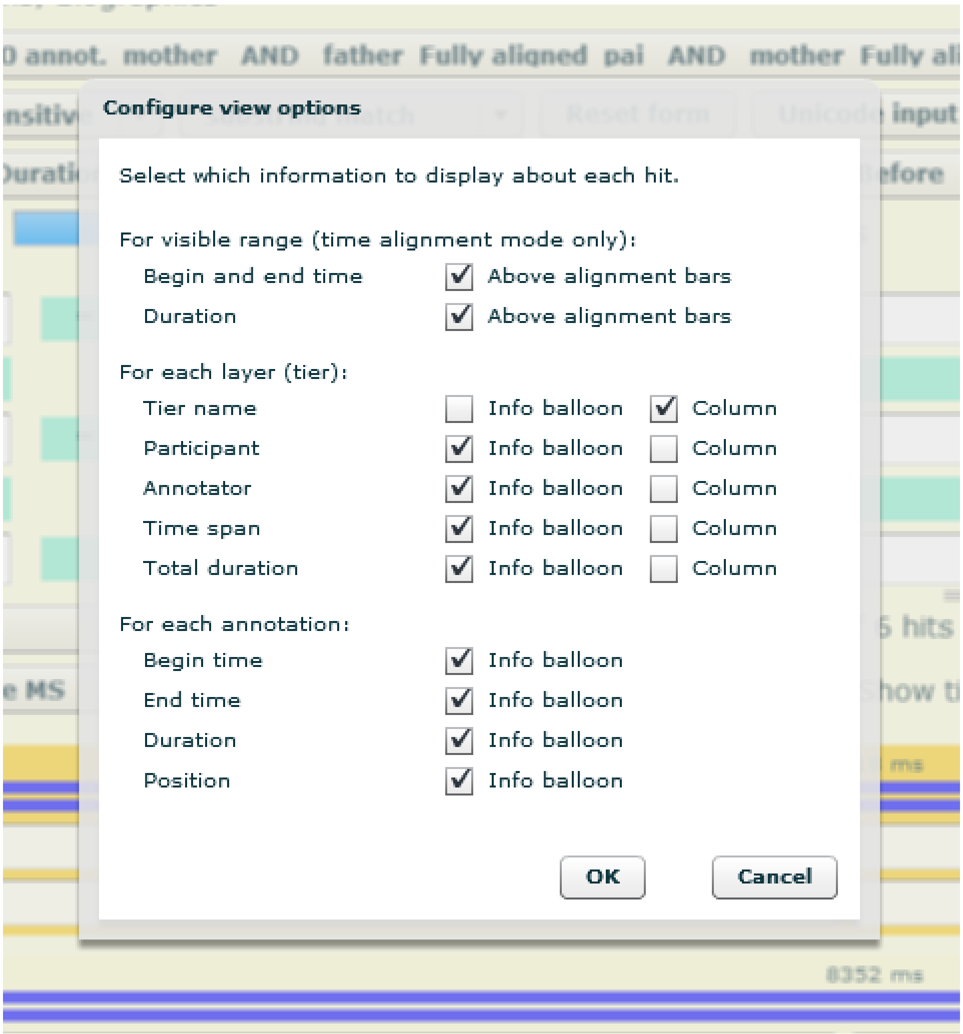
anoiteceu, no dia seguinte estamos o derramando — SmngP@045

00:03:33.087	10 seconds	00:03:43.087

it is drying for two days — SmngE@045

Figure 1.13. Time Alignment - adjusted scale

Finally, on the left of the option **Show Time Alignment**, you can click on **Info View Options** and decide what has to be shown or hidden in the balloon appearing when hovering over the results, in the purple column next to the results, or above the results (i.e. the time options), simply by checking or unchecking the (un)required information. See figure below.



Configure view options

Select which information to display about each hit.

For visible range (time alignment mode only):

Begin and end time	☑ Above alignment bars	
Duration	☑ Above alignment bars	

For each layer (tier):

Tier name	☐ Info balloon	☑ Column
Participant	☑ Info balloon	☐ Column
Annotator	☑ Info balloon	☐ Column
Time span	☑ Info balloon	☐ Column
Total duration	☑ Info balloon	☐ Column

For each annotation:

Begin time	☑ Info balloon
End time	☑ Info balloon
Duration	☑ Info balloon
Position	☑ Info balloon

Figure 1.14. Info View Options

Search hints

If you would like to use both Exact match and Substring match in one query, use the Regular expressionoption. In places where you would like to have an exact match use the ^ and $ signs to match the beginning and the end of a string (e.g. ^of$), otherwise just enter a word for the substring match.

Wild card. Instead of using the # as in the Single Layer Search, you can use the regular expression .+ to indicate any character (the dot) one or more times (the plus sign). The NOT(...) construction on the other hand can be used in the Multiple Layer Search in the same way as described in section 1.2.

One final, but not less important, remark concerns the placing of relatively restrictive search strings. As we saw earlier, the hierarchy of the rows in the query does not reflect the hierarchy in the data. While this is perfectly true, we advise you to place restrictive search strings in the left most field on in the upper most row possible and the least restrictive search string in the right most field of the lowest row possible. The reason for this is the order in which the search engine considers the search strings in the query. If it finds a restrictive search string it can filter out all the other possibilities, but if it finds a less restrictive search string it has to consider all the matches of this search string. Because of this, the search might take much more time if the non-restrictive strings are placed before the restrictive ones.

1.4. Regular Expressions

Regular expressions allow users to create complicated queries. Below follows a list of most commonly used regular expressions together with explanations and some potential uses.

- [abc] means "a or b or c", e.g. query "[br]ang" will match both "adbarnirrang" and "bang"
- [^abc] means "begins with any character but a,b,c", e.g. query [^aeou]ang will match "rang" but not "baang"
- [a-zA-Z] means "a charater from a/A through z/Z", e.g. b[a-zA-Z] will match "bang", "bLang" or "baang" but not "b8ng"
- . (the dot) means "any character", e.g. "b.ng" will match "bang", "b8ng", but not "baang"
- X* means "X zero or more times", e.g. "ba*ng" will match "bng", "bang", "baang", "baaang" etc.
- X+ means "X one or more time", e.g. "ba+ng" will match "bang", "baang" but not "bng"
- ^ means "the beginning of the annotation", e.g. "^ng" will match "ngabi" but not "bukung"
- $ means "the end of the annotation", e.g. "ung$" will match "bukung" but not "ngabi"

  Examples

  - ^[pbtd][^aeiou]

    You can use this expression to search for complex onsets. It will find words that start with one of the plosives ("p","b","t","d") followed by a character that is not a vowel ("a","e","i","o","u"). An example of a matching word is "tsakeha"

  - [^n]g$

    You can use this expression in case you want to search for annotations ending with a "g", but not with "ng". In Dutch, you will find "snelweg" and "maandag" as the results but not words as "bang".

  - ^k.+k$

    You can use this expression if you want to search for annotation starting and ending with "k" and with one or more character between them, e.g. "kitik" or "kanak-kanak"

- ^(.+)\1$

  You can use this expression to search for words that are reduplicated. When you put something in bracketes, you create a variable (.+), which you can refer to as "\1". This expression then searches for an annotation that starts with one or more random characters followed by that same sequence of characters. This expression will match for instance "kulukulu".

More about regular expressions...

The following tables have been created by a user of ELAN (an annotation tool which has the same search mechanism as TROVA). They may result quite useful also for other users since they offer a simple and clear overview of the main symbols (partly different from the ones just seen) used in regular expressions, with a short explanation and an example for each of them. Bear in mind that the examples are taken from the language that the user is being researching, so do not pay attention to the meaning of the words but to the working mechanism of the regular expressions.

Table 1.1. Symbols

Symbols	Place	Meaning
\b	at the beginning and/or end of a string	word boundary
\w+	at the end of a string	variable end of word
.	anywhere	any letter
.*	between spaces	any string of letters between spaces/any word
.*\	between spaces	any string of words
(x|y)	anywhere	either x or y
[^x]	place at the beginning	not x
(....)\l	anywhere	words with four reduplicated letters
?	after a letter	the preceding letter is optional

Table 1.2. Search for particular word forms (examples)

Symbols	Hits	Examples
sa	all words containing the string *sa*	*sa*, *vasaku*, *sahata*, *tisa*
\bsa	all words starting with *sa*	*sa*, *sahata*, *sana*; NOT vasaku, tisa
\bsa\b	all words *sa*	*sa*
\bsa..\b	all words consisting of *sa* + two letters that follow *sa*	*saka*, *saku*, *sana*
\bsa\w+	all words beginning with *sa*, but not the word *sa* by itself	*sahata*, *sana*
\b.*ana\b	al words ending in *ana*	*sinana*, *tamuana*, *sana*, *bana*, *maana*
(....)\l	all words with four reduplicated letters	*pakupaku*, *vapakupaku*, *mahumahun*, *vamahumahun*
\b(....)\l	all words beginning with four reduplicated letters	*pakupaku*; NOT vapakupaku
\b(....)\lana\b	all words beginning with four reduplicated letters and ending in *ana*	*vasuvasuana*, *hunuhunuana*
\bva(....)\l	all words consisting of the prefix va- + four reduplicated letters	*vapakupaku*, *vagunagunaha*

Symbols	Hits	Examples
\bvahaa?\b	all tokens of vahaa and vaha	*vahaa* and *vaha*

Table 1.3. Search for particular sequences of words (examples)

Symbols	Hits	Examples
\bsaka\b .* \bhaa	string of 3 words: (1) *saka*; (2) any word; (3) the word *haa* by itself or with suffixes	*saka antee haa*; *saka abana haari*; *saka kabuu haana*
saka .* \bhaa\w+	string of 3 words: (1) *saka*; (2) any word; (3) a word beginning with *haa*, but NOT the word *haa* by itself	*saka abana haari*; *saka kabuu haana*
(\bsaka\b|\bsa\b) \bpaku\b	2-word string consisting of *saka* or *sa* and *paku*	*saka paku*; *sa paku*
(\bsaka\b|\bsa\b) .* \bvaha\b	strings of 3 words: (1) *saka* or *sa*; (2) any word; (3) *vaha*	*saka tii vaha*; *sa tapaku vaha*
(\bsaka\b|\bsa\b) (....)\l \bhaa	strings of 3 words: (1) *saka* or *sa*; (2) any word with four reduplicated letters; (3) the word *haa* or a word beginning with *haa*	*sa natanata haa*; *saka natanata haana*