

Recherche par ligne de commande dans ELAN-CorpA et ELANWebSearch

Principe: Dans le menu *Search* de [ELAN-CorpA](#), l'item *multiple files search* permet de rechercher par ligne de commande quelque chose (une séquence) quelque part (dans une tier). La requête s'écrit simplement sous la forme [tier=séquence].

La recherche peut se faire sur une tier ou sur un ensemble de tiers partageant le même *type*. Pour distinguer le nom d'une tier du nom d'un *type* de tier, on mettra le signe \$ devant le nom d'une tier et rien devant le nom d'un *type* (on privilégie la recherche sur un ensemble de tier).

ex: [\$mb=-i] pour rechercher (le morphème) -i dans la tier intitulé mb

[mb=-i] pour rechercher (le morphème) -i dans toutes les tiers de *type* mb (mb@SP1, mb@SP2...)

les *expressions régulières* sont acceptées après le signe =

ex: [mb=he:|ha:j|hi:] pour rechercher les 3 formes he:, ha:j ou hi: dans les tiers de type mb

Pour rechercher une étiquette morpho-syntaxique simple comme *N* ou *V...*, on l'encadrera de codes \b (début de mot) pour éviter le retour de toutes les étiquettes comportant ces lettres. Les caractères séparateurs (. ou \) utilisés entre les étiquettes des annotations complexes (ex: *DET.DEICT*) seront considérés comme des espaces dans une expression régulière comme \bN\b et seules les annotations comportant un N isolé seront retournées.

Contraintes verticales

A l'intérieur d'une paire de crochets [] on peut mettre des contraintes sur les gloses dépendantes verticalement.

Le symbole

& représente une contrainte de type '*fully aligned*' (les deux annotations sont strictement alignées)

< représente une contrainte de type '*inside*' (la partie gauche est à l'intérieur ou égale à la partie droite)

> représente une contrainte de type '*includes*' (la partie gauche contient ou est égale à la partie droite)

ex: [mb = -í: & ge = DIST] renvoie les occurrences du morphème -í: dont la glose est DIST

| ndwátsə ménvərvàn yá:í: burúk kəŋ á məs / (SAY_BC_INT_02_Burial_SP2_44) | | | | | |
|---|-----------|---------|-----------|-----------|------------------------------|
| ndwátsə | ménvərvàn | yá:í: | burúk kəŋ | á | məs / |
| ndwátsə | ménvər | = wàn | ya:í | -í: | burúk kəŋ á məs / |
| old_person | grownup | 2PL.POS | earth | DIST | all COP2 3SG.AOR die / |
| ADJ | N | PRO | N | DET.DEICT | ADV PTCL.SYNT.hom PN.TAM V / |
| the oldest man of the whole country that has died, | | | | | |

fig. 1 [mb=-í: & ge=DIST]

Le résultat d'une requête listera toutes les annotations contenant la séquence recherchée, **ainsi que les valeurs des éléments de contrainte**.

Ainsi, si on veut récupérer la valeur d'une glose associée à l'annotation recherchée, on ajoutera la tier de cette glose comme contrainte avec la valeur point (.). Le *point* ayant valeur de '*quelque soit*', c'est une contrainte qui en fait n'en est pas une mais qui va permettre de retourner la valeur actuelle de cette autre tier dans le résultat de la requête.

ex: [rx = -í: & ge = .] retrouve les occurrences du morphème -í: et renvoie leur glose (ge) associée

ex: [rx = PTCL & ge = . < mot = .] retrouve chaque occurrence de *particule* (rx=PTCL) et renvoie la valeur de sa glose (ge) et le mot dont elle est issue (< parce que les éléments *ge* et *rx* sont des annotations de morphèmes *mb*, eux-mêmes inclus dans des éléments *mot*)

SAY_BC_INT_01_RELIGION.EAF :

- #1 [PTCL.NEG] #2 [NEG1] #3 [bà] { SAY_BC_INT_01_Religion_SP1_006 }
- #1 [PTCL.NEG] #2 [NEG2] #3 [hǎ] { SAY_BC_INT_01_Religion_SP1_008 }
- #1 [PTCL.SYNT] #2 [COP1] #3 [nə] { SAY_BC_INT_01_Religion_SP1_010 }
- #1 [PTCL.TOP] #2 [indeed] #3 [kàm] { SAY_BC_INT_01_Religion_SP1_014 }
- #1 [PTCL] #2 [DM] #3 [tò:] { SAY_BC_INT_01_Religion_SP1_014 }
- #1 [PTCL.NEG] #2 [NEG1] #3 [bà] { SAY_BC_INT_01_Religion_SP1_016 }
- #1 [PTCL.NEG] #2 [NEG2] #3 [hǎ] { SAY_BC_INT_01_Religion_SP1_016 }
- #1 [PTCL.SYNT] #2 [COP1] #3 [nə] { SAY_BC_INT_01_Religion_SP1_019 }
- #1 [PTCL.SYNT] #2 [COP1] #3 [nə] { SAY_BC_INT_01_Religion_SP1_021 }
- #1 [PTCL.NEG] #2 [NEG2] #3 [hǎ] { SAY_BC_INT_01_Religion_SP1_026 }

| | | | | | | |
|--|------|--------------|----|-----------|--------|-------------|
| tò: àmá mètàyá: fi wárès kàm // (SAY_BC_INT_01_Religion_SP1_014) | | | | | | |
| tò: | àmá: | mètàyá: | fi | wárès | kàm | // |
| tò: | àmá: | mètàyá: | fi | wâr -ès | kàm | // |
| DM | but | 1PL.REM.ICPL | do | fetishDEF | indeed | // |
| PTCL | CONJ | PN.TAM | V | N | ART | PTCL.TOP // |

fig. 2 [rx=PTCL & ge=. < mot=.]

ex: [mb=ǝim < mot=.] renvoie toutes les formes fléchies (mot) du morphème ǝim (mb)

- SAY_BC_INT_01_RELIGION.EAF :
- #1 [ǝim] #2 [ǝim] { SAY_BC_INT_01_Religion_SP1_188 }
- #1 [ǝim] #2 [ǝimtə] { SAY_BC_INT_01_Religion_SP1_201 }
- #1 [ǝim] #2 [ǝim] { SAY_BC_INT_01_Religion_SP1_328 }
- #1 [ǝim] #2 [ǝimni] { SAY_BC_INT_01_Religion_SP1_330 }

| | | | | | | |
|--|--------------|-----------|----------|-------|-------|---------------|
| àmá: mètàyá: ǝimni á vi: kúrúm // (SAY_BC_INT_01_Religion_SP1_330) | | | | | | |
| àmá: | mètàyá: | ǝimni | á | vi: | kúrúm | // |
| àmá: | mètàyá: | ǝim-ni | ká | vi: | kúrúm | // |
| but | 1PL.REM.ICPL | call INCH | at | mouth | only | // |
| CONJ | PN.TAM | V | PTCL.EXT | PREP | N | ADV/BW.HAU // |

fig. 3 [mb=ǝim < mot=.]

Contraintes horizontales

Pour rechercher une annotation suivie d'une autre annotation, on écrira successivement les 2 paires de crochets [] correspondant aux annotations recherchées, séparées par une paire d'accolades { } dans laquelle seront indiquées les contraintes de distances entre ces annotations.

Les opérateurs de contrainte horizontale sont :

=, <, > qui permettent de rechercher respectivement une annotation à une distance égale (=), inférieure (<) ou supérieure (>) à la valeur suivant le symbole.

ex: {mot < 3}

signifie que le mot de droite doit être à une distance de un ou deux mots du mot de gauche.

Si les recherches portent sur des annotations comportant chacune des contraintes verticales, les contraintes horizontales doivent être définies pour chaque tier (ou type) apparaissant dans les 2 séries de crochets.

ex: [rx=DET & mb=.] [rx=1 & mb=1] [rx=\bN\b & mb=.]

renvoie les morphèmes (mb=.) de type *déterminant* (rx=DET) directement (mb=1) suivi d'un morphème (mb=.) de type *nom* (rx=N)

ex: [rx=DET < mot=.] [rx<5 & mot=1] [rx=\bN\b <mot=.]

renvoie les mots (mot=.) comportant un déterminant (rx=DET) et qui sont directement (mot=1) suivis d'un nom (rx=N).

'< mot=' signifie que le déterminant (rx=DET) peut être un constituant (affixe) du mot.

{rx<5 & mot=1} : le déterminant et le nom doivent apparaître dans 2 mots qui se suivent (mot=1) mais il faut aussi donner une contrainte sur rx parce que rx apparaît dans les 2 paires de crochets. La valeur 5 est prise par défaut, en supposant qu'un mot ne comportera pas plus de 5 morphèmes.

| • SAY_BC_INT_01_RELIGION.EAF : □ | | | |
|---|---|--|----------------------------------|
| □ #1 [DET] [PN.TAM] #2 [wón] [tá] { SAY BC INT 01 Religion | | tô: ka yel gya: wús táyi mbwa:ya // | (SAY_BC_INT_01_Religion_SP1_095) |
| □ #1 [DET] [N] #2 [gyá:] [wús] { SAY BC INT 01 Religion SP1 | | tô: ka yel gya: wús táyi mbwa:ka // | |
| □ #1 [DET] [N] #2 [gyá:] [wús] { SAY BC INT 01 Religion SP1 | ▶ | tô: ka yeli gyá: wús táyi mbwa: =kə // | |
| □ #1 [DET.DEICT] [PN.TAM] #2 [-i:] [yá:] { SAY BC INT 01 | | DM 2SG.FUT see PL boil 3PL.FUT.ITER appear2SG.OBJ // | |
| □ #1 [DET] [N.P] #2 [gyá:] [gánám] { SAY BC INT 01 Religion | | PTCL PN.TAM V DET N PN.TAM V PRO // | |
| □ #1 [DET] [N.P] #2 [gyá:] [kú:bá] { SAY BC INT 01 Religion | | | |

fig. 4 [rx=DET < mot=.] [rx<5 & mot=1] [rx=N <mot=.]

ex: [rx=V < mot=.] [rx<5 & mot=0] [rx=PRO < mot=.]

renvoie les mots (mot=.) qui sont des verbes (rx=V) et qui contiennent {mot=0} des pronoms (rx=PRO)

Il s'agit ici de 2 annotations successives à l'intérieur d'un même mot.

Il est possible d'utiliser le symbole ? (des expressions régulières) après une paire de crochets pour signifier la présence non obligatoire du contenu de l'annotation recherchée.

ex: [rx=DET < mot=.] [rx<5 & mot=1] [rx=ADJ < mot=.] ? [rx<5 & mot=1] [rx=\bN\b < mot=.]

recherche un *déterminant* éventuellement (?) suivi d'un *adjectif* puis d'un *nom*.

Pour écrire une telle requête, on commencera par écrire ce que l'on recherche (en considérant que les mots peuvent contenir jusqu'à 3 morphèmes):

[rx=DET] [rx<3] [rx=ADJ] ? [rx<3] [rx=\bN\b]

pour contraindre les mots à être successifs, on doit ajouter la contrainte {mot=1} entre les annotations

[rx=DET] [rx<5 & mot=1] [rx=ADJ] ? [rx<5 & mot=1] [rx=\bN\b]

pour éviter une erreur de syntaxe, les éléments de contraintes doivent apparaître dans chaque groupe d'annotations recherchés d'où

[rx=DET < mot=.] [rx<5 & mot=1] [rx=ADJ < mot=.] ? [rx<5 & mot=1] [rx=\bN\b < mot=.]

La requête suivante permet une certaine distance (1 ou 2) entre les différents mots

[rx=DET < mot=.] [rx<10 & mot<3] [rx=ADJ < mot=.] ? [rx<10 & mot<3] [rx=\bN\b < mot=.]

Recherche de la x^{ième} occurrence

Lorsqu'on fait une recherche de co-occurrences avec comme contrainte horizontale {tier < x} toutes les occurrences de la deuxième annotation à la distance 0 à x-1 de la première sont renvoyées. Pour rechercher uniquement la première, seconde ou x^{ième} occurrence d'une annotation donnée (après une précédente annotation), on rajoutera le symbole : suivi d'un chiffre (1 pour la première occurrence, 2 pour la deuxième...) précisant la position de l'annotation suivante après la précédente.

ex: [rx=\bV\b < mot=.] [rx<10 & mot<6:1] [rx=\bN\b < mot=.]

renvoie les verbes suivi du premier nom à une distance de 5 mots maximum

ex: [rx=\bV\b < mot=.] [rx<10 & mot<3:1] [rx=\bN\b < mot=.] [rx<10 & mot<3:1] [rx=\bN\b < mot=.]

recherche les verbes suivis de 2 noms distants les uns des autres d'au plus 2 mots.

Pour que [V][N][N] soit en plus dans la même unité prosodique, il faudra rajouter une contrainte horizontale sur ref et pour cela il faut ajouter les contraintes <ref=. dans chaque paire de crochets:

[rx=\bV\b < mot=. <ref=.] [rx<10 & mot<3:1 & ref=0] [rx=\bN\b < mot=. <ref=.]
[rx<10 & mot<3:1 & ref=0] [rx=\bN\b < mot=. <ref=.]

Recherche de 2 morphèmes à l'intérieur d'un même mot (mot=0)

[rx=PNG < mot=.] [rx<3 & mot=0] [rx=TAM < mot=.]

| | |
|----------------------------|--|
| #1 PNG- -TAM | #2 ba:ka:je: ba:ka:je: |
| #1 PNG -TAM | #2 takè: takè: |
| BEJ_MV_NARR_01_shelter_133 | |
| w?ara:wo: ba:'ka:j / | |
| w?ara:wo: | ba:ka:je: / |
| w= | ?ara:w =o: ba= a- ka -aj =e: / |
| DEF.SG.M= | friend -POSS.3SG, NEG.O, 1SG- be -OPT -REL . |
| DET= | N -PRO PTCL= PNG- V1.IRG -TAM -CONJ . |

Recherche d'un motif avec mise en facteur (simplification d'écriture) :

recherche d'un mot constitué d'un radical N et du premier suffixe PRO à une distance inférieure à 5.

mot=0: [rx=\bN\b]{rx<5:1}[rx=PRO]

- #1 |N| |=PRO| #2 |w?ara:wo:| |w?ara:wo:|
- #1 |N.F| |=PRO| #2 |tifa:to:n| |tifa:to:n|
- #1 |SBJ.N| |=PRO| #2 |?ara:wil| |?ara:wil|
- #1 |N.M| |=PRO| #2 |imhi:ni| |imhi:ni|
- #1 |N.M| |=PRO| #2 |e:gaq?aje:| |e:gaq?aje:|
- #1 |N.M| |=PRO| #2 |ama:ro:n| |ama:ro:n|
- #1 |N.M| |=PRO| #2 |sama:ro:n| |sama:ro:n|

Recherche de N suivi de P à une distance < 5 dans la même phrase

[rx=N < tx=.]{rx<5 & tx=0} [rx=P < tx=.] ou bien tx=0:[rx=N]{rx<5} [rx=P]

- #1 |SBJ.N| |=PRO| #2 |'ka:me /| |'ka:me /|
- #1 |NP| |=POSTP| #2 |i'fergib i:'fe:...| |i'fergib i:'fe:...|
- #1 |NP| |TAM.PNG-| #2 |i'fergib i:'fe:...| |i'fergib i:'fe:...|
- #1 |N| |=PRO| #2 |ka:'mji 'harawa...| |ka:'mji 'harawa...|
- #1 |N.M| |POSTP| #2 |'biridhe: ?e'ja...| |'biridhe: ?e'ja...|
- #1 |N.M| |=PRO| #2 |'biridhe: ?e'ja...| |'biridhe: ?e'ja...|
- #1 |N.M| |PNG-| #2 |'biridhe: ?e'ja...| |'biridhe: ?e'ja...|

Recherche de la première occurrence (:1) d'un N suivi d'un P dans la même phrase

[rx=N < tx=.]{rx<5:1 & tx=0} [rx=P < tx=.] ou bien tx=0:[rx=N]{rx<5:1} [rx=P]

- #1 |SBJ.N| |=PRO| #2 |'ka:me /| |'ka:me /|
- #1 |NP| |=POSTP| #2 |i'fergib i:'fe:...| |i'fergib i:'fe:...|
- #1 |N| |=PRO| #2 |ka:'mji 'harawa...| |ka:'mji 'harawa...|
- #1 |N.M| |POSTP| #2 |'biridhe: ?e'ja...| |'biridhe: ?e'ja...|
- #1 |CN.M| |POSTP| #2 |li:l'?'awgeb //| |li:l'?'awgeb //|
- #1 |TAM.PNG-| |-PNG| #2 |kir'ra:bikti:n ...| |kir'ra:bikti:n ...|

Recherche de N.M suivi de PRO ou bien de CONJ dans la même phrase

[rx=\bN.M\b < tx=.]{rx=1 & tx=0}[rx=PRO|CONJ < tx=.] ou bien

tx=0:[rx=\bN.M\b]{rx=1} [rx=PRO|CONJ]

- #1 |N.M| |=PRO| #2 |adit w'hisi gaj...| |adit w'hisi gaj...|
- #1 |N.M| |=CONJ| #2 |'b?adaqwa 'k'wo:...| |'b?adaqwa 'k'wo:...|
- #1 |N.M| |=CONJ| #2 |'b?adaqwa 'k'wo:...| |'b?adaqwa 'k'wo:...|
- #1 |N.M| |=PRO| #2 |'agari'?'andiga'...| |'agari'?'andiga'...|
- #1 |N.M| |CONJ| #2 |'ji?o:'biri 'na...| |'ji?o:'biri 'na...|

Présence éventuelle (?) d'un élément :

Recherche d'un verbe suivi ou pas d'un TAM puis d'un verbe.

[rx=V]{rx=1}[rx=TAM]?{rx=1}[rx=V]

#1 |V1| |PNG.TAM| |N.VO| ||

#1 |V1| |PNG.TAM| |V1.AUX| ||

#1 |V1| |PNG.TAM| |V1| ||

#1 |V1| |PNG.TAM| |V2| ||

#1 |V2| |PNG.TAM| |ADV| ||

#1 |V2| |-der.V2| ||

#1 |V1.AUX| |V1| ||

#1 |V1.AUX| |N.VO| ||