

Útmutató az adatbázisok használatához

1. Az adatbázis szövegei a böngészőben

A böngészőben az uráli szövegek UTF-8 karakterkódolással jeleníthetők meg hibátlanul. Az UTF-8 karakterkódolást a böngésző legfelső sávjában állíthatjuk be:

Nézet (View) > Karakterkódolás (Text Encoding) > UTF-8 vagy Unicode

A *Nézet* és *Karakterkódolás* menüpontok elnevezései böngészőtől függően kissé eltérhetnek az itt megadottól.

2. Az adatbázis szövegei a böngészőn kívül

Ahhoz, hogy a számítógépen megfelelően jelenjenek meg az uráli szövegek, a Charis SIL fontkészlet telepítése szükséges.

Ezt innen tölthetjük le: <https://software.sil.org/charis/> (*Download > Download current version*)

Az új fontkészlet telepítésének módja attól függ, hogy milyen operációs rendszerünk van. Windows esetében részletes útmutatást találunk ezen a linken:

<https://support.microsoft.com/hu-hu/help/314960/how-to-install-or-remove-a-font-in-windows>

Az adatbázis nyelvészeti kutatásra akkor válik a leginkább alkalmassá, ha az egyes szövegrészek alatt látjuk a hozzájuk tartozó, különféle szintű elemzéseket is. Ezt a nézetet, valamint az adatbázis legjobb kereshetőségét az ELAN programcsomaggal érjük el.

Az ELAN innen tölthető le: <https://tla.mpi.nl/tools/tla-tools/elan/>

A weboldalon válasszuk a *Download the latest version* menüpontot, majd a betöltő oldalon keressük meg az operációs rendszerünknek megfelelő ELAN telepítőt. Windows esetén célszerű a legelső lehetőséget választani, amely a megfelelő Java környezetet is tartalmazza.

Az adatbázis használatát megkönnyíti az, ha megismerkedünk az ELAN-ban használt alapvető fogalmakkal. A használati útmutató rövid változata itt érhető el:

https://www.mpi.nl/corpus/html/elan_ug/index.html

Az ELAN telepítése után lehetőségünk van az adatbázis egy vagy több szövegének betöltésére, és a szövegekben egyszerű vagy többszintű, komplex keresésekre.

2.1. Szöveg betöltése az ELAN-ba

1. Töltsük le a választott .eaf fájlokat.

Ha például új színjai hanti szövegeket szeretnénk vizsgálni:

Adatbázisok > Színjai hanti > Új szövegek, 4. ELAN fájlok, .eaf fájlok > a választott .eaf fájlok letöltése a számítógép tetszőleges mappájába

Ezt úgy tehetjük meg, hogy a betöltő, *Index of...* című oldalon jobb egérgombbal klikkelünk a fájl nevére, és a *Save as... / Mentés másként...* opciót választjuk.

2. Ha hanganyag is tartozik a szövegekhez, le lehet tölteni, de nem feltétlenül szükséges.

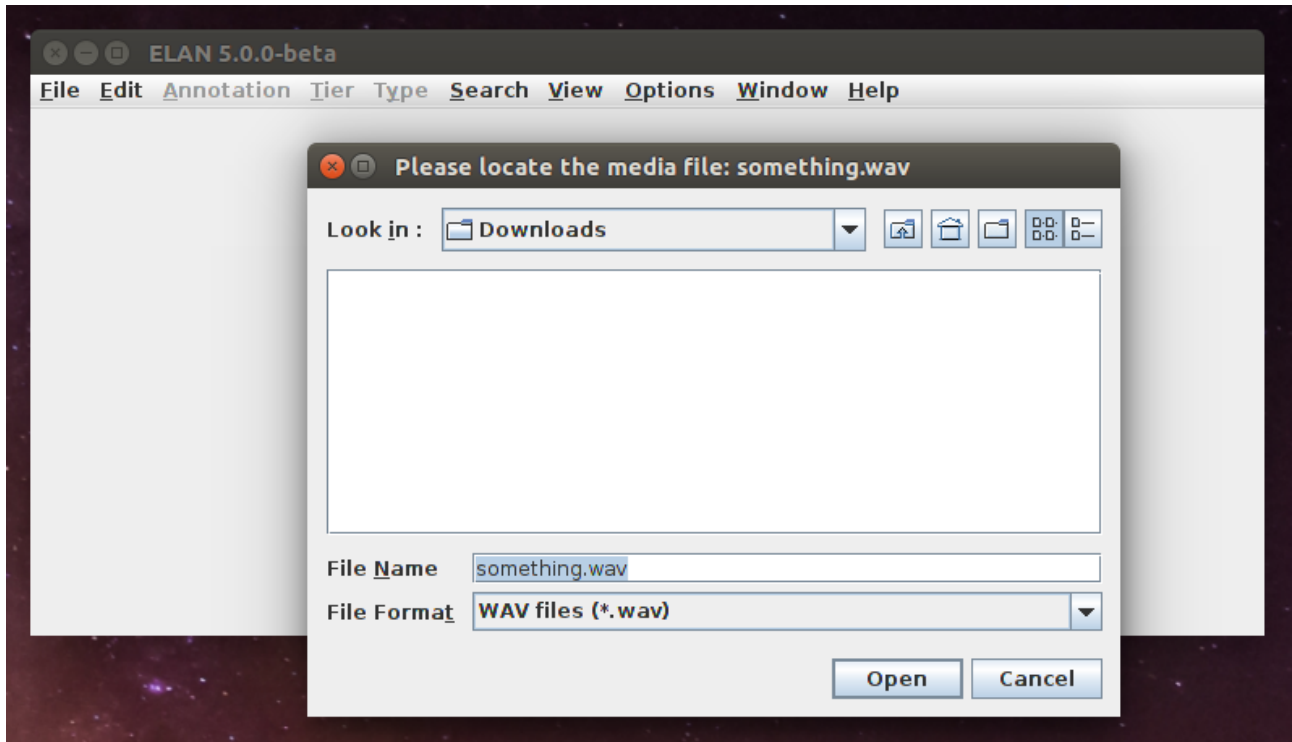
Például az új színjai hanti szövegek esetében:

Adatbázisok > Szinjai hanti > Új szövegek, 4. ELAN fájlok, audiofájlok > a választott .wma vagy .wav fájlok letöltése a számítógép tetszőleges mappájába

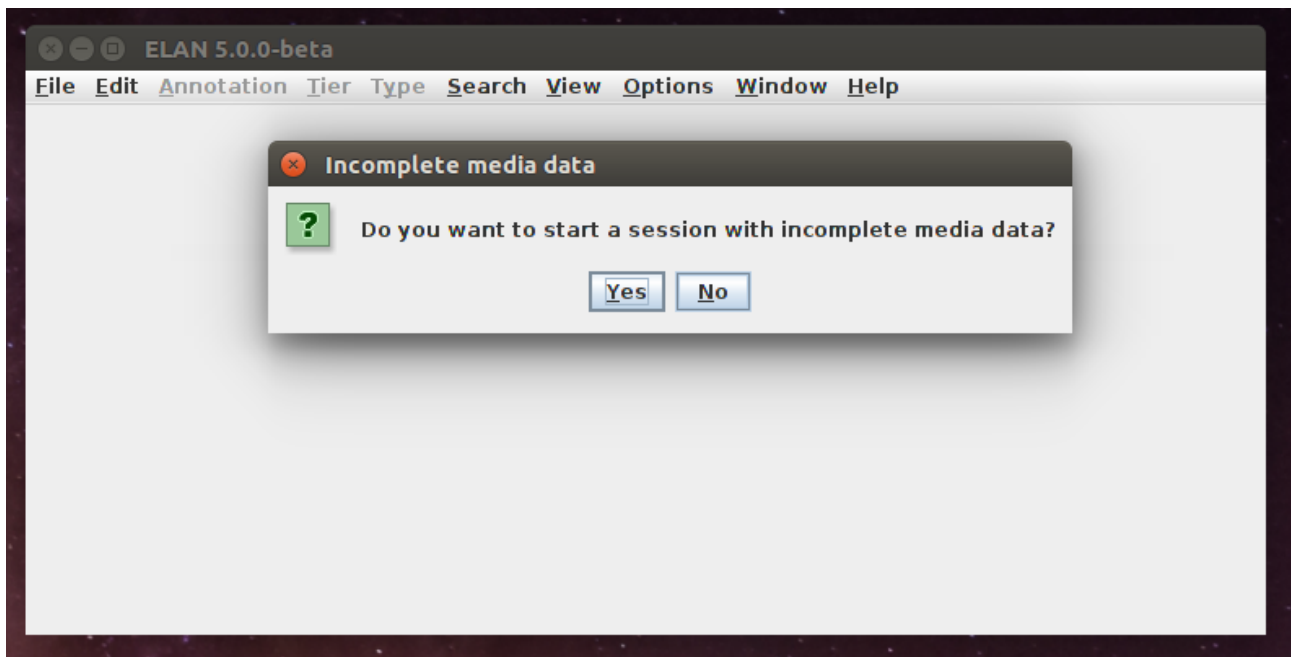
3. Megnyitjuk az ELAN-t, és a legfelső menüsávban megnyitjuk a fájlt:

File > Open... > kiválasztunk egy letöltött .eaf fájlt > *Open*

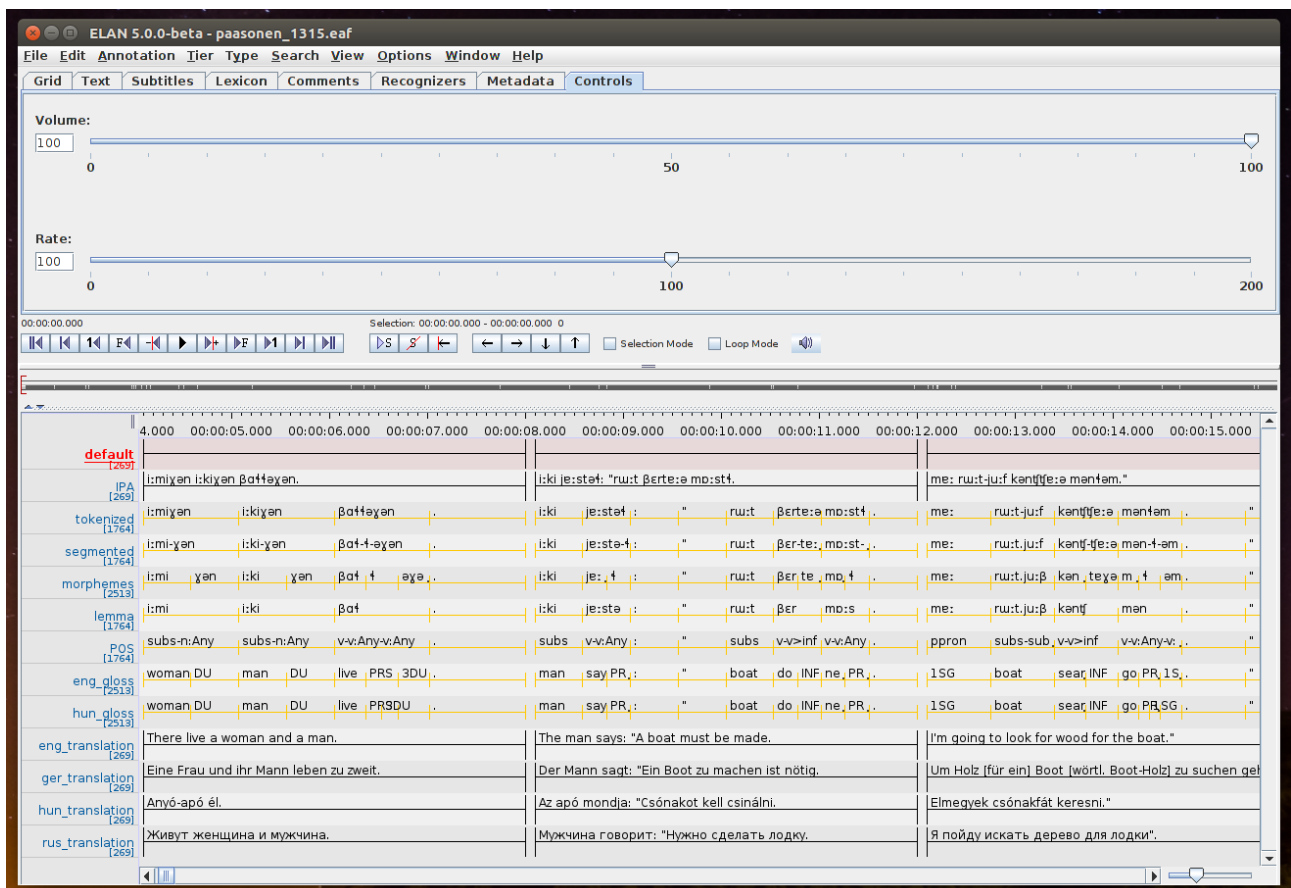
Ekkor megjelenik egy ablak, ahol megadhatjuk, hogy tartozik-e hanganyag az adott fájlhoz:



Ha nincs hanganyag, vagy csak nem akarjuk betölteni, válasszuk a *Cancel* opciót. Ekkor megjelenik egy ablak, ahol a *Yes* megnyomásával megerősítjük, hogy nem akarunk hanganyagot betölteni:

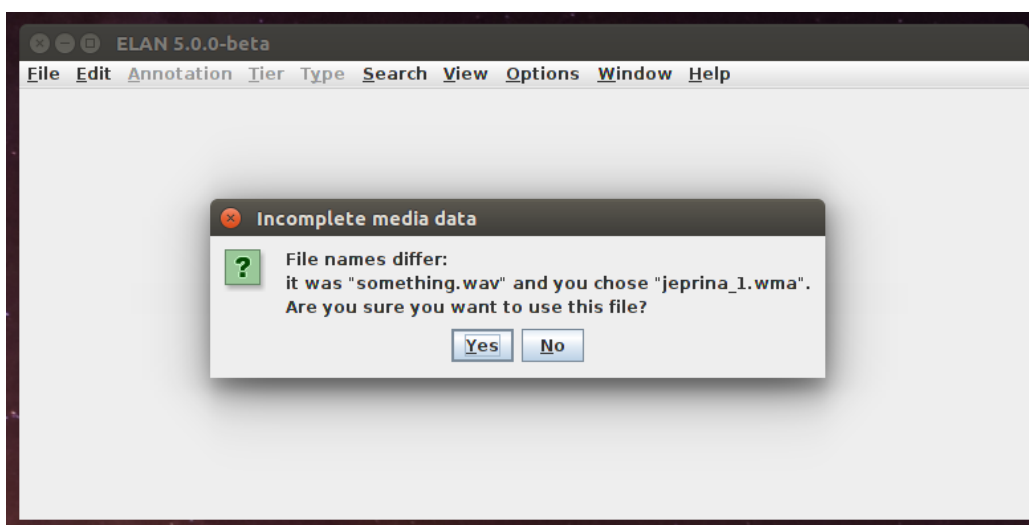


Ezután betölt a választott szövegfájl. Az ablak alsó görgetősávjával tudunk előre és hátra mozogni a szövegben, a jobb alsó sarokban található csúszkával nagyíthatjuk és kicsinyíthetjük a szöveget:

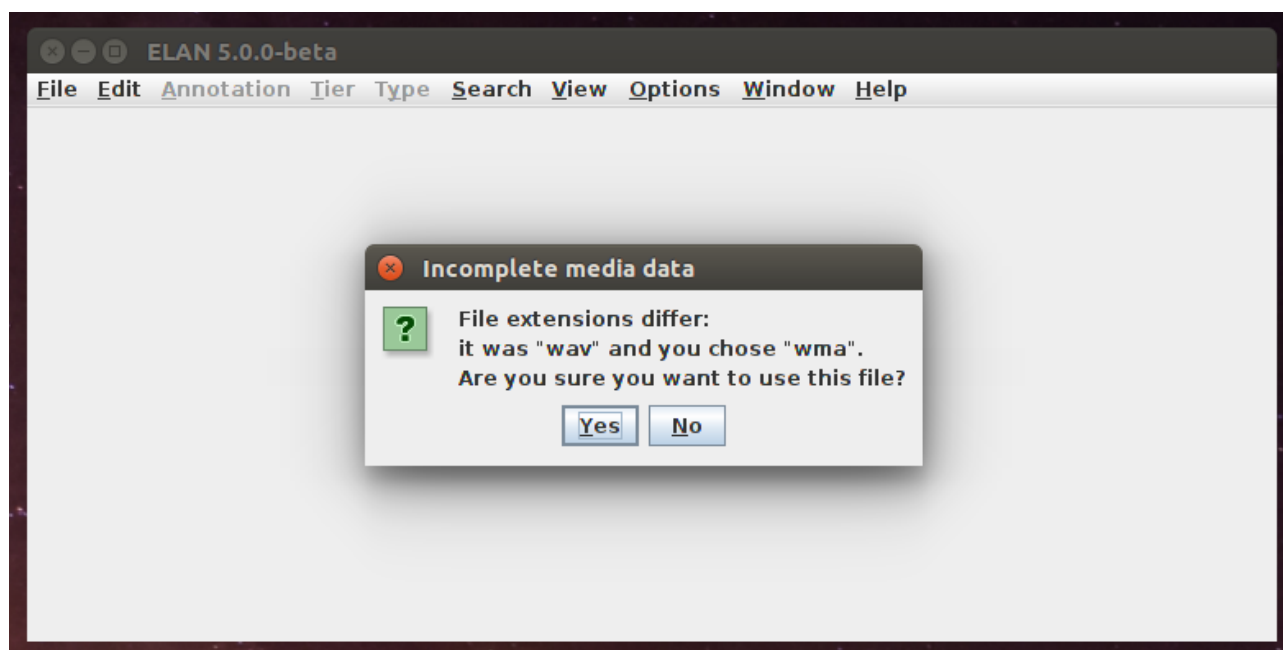


Ha szeretnénk hanganyagot is betölteni, akkor az .eaf fájl kiválasztása után megjelenő ablakban keressük meg a weboldarról letöltött hangfájlt. Amennyiben a hangfájl nem .wav típusú (például a korábbi példában .wma a kiterjesztés), akkor a *File Format* legördülő menüjében a *WAV files*

(**.wav*)-ot állítsuk át *All Files*-ra. Ezután válasszuk ki a megfelelő hangfájlt, és nyomjuk meg az *Open* gombot. Ekkor megjelenik egy felugró ablak, ahol a *Yes* megnyomásával erősítjük meg, hogy ezt a hangfájlt szeretnénk használni:



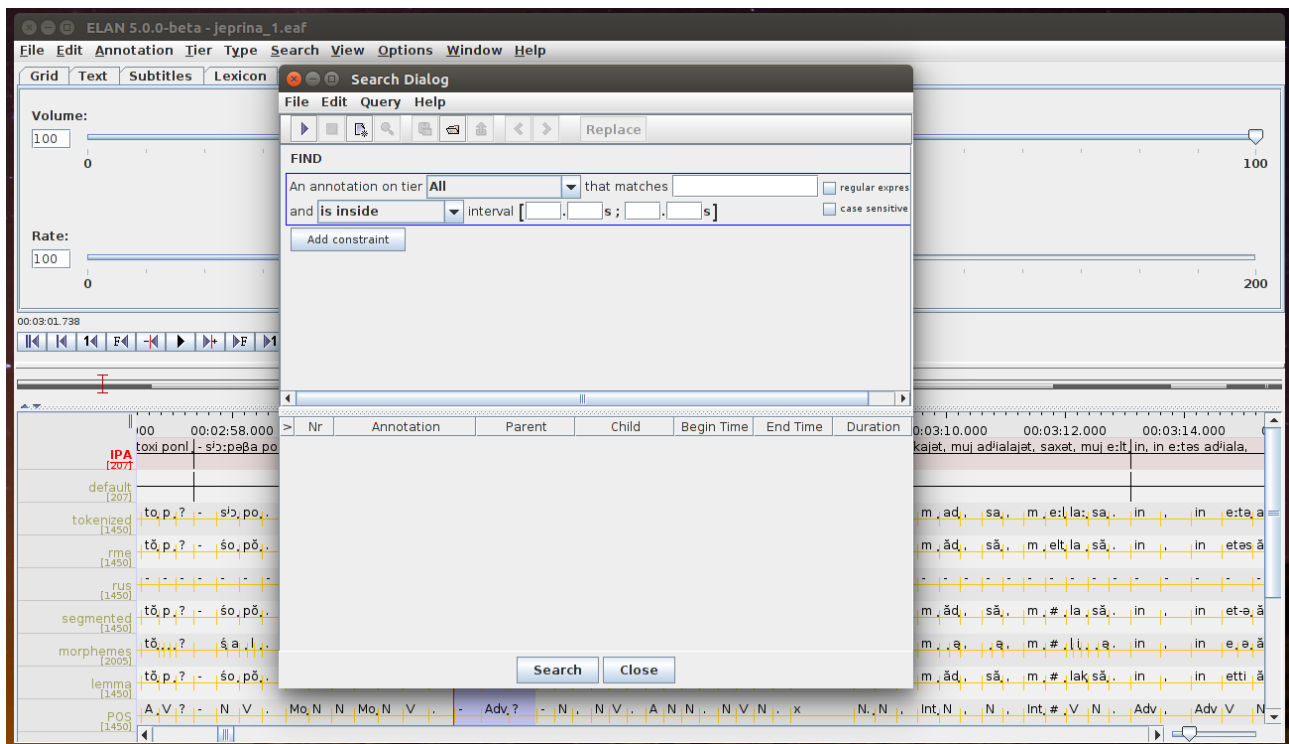
Amennyiben a fájl nem *.wav* kiterjesztésű, azt is meg kell erősítenünk egy *Yes*-szel a következő ablaknál:



Ekkor betölt a választott szövegfájl, amihez ezúttal hanganyagot is le tudunk játszani a középső sávban látható vezérlők segítségével, vagy egyszerűen a *Ctrl+Space* használatával.

2.2. Egy szövegen belüli keresés

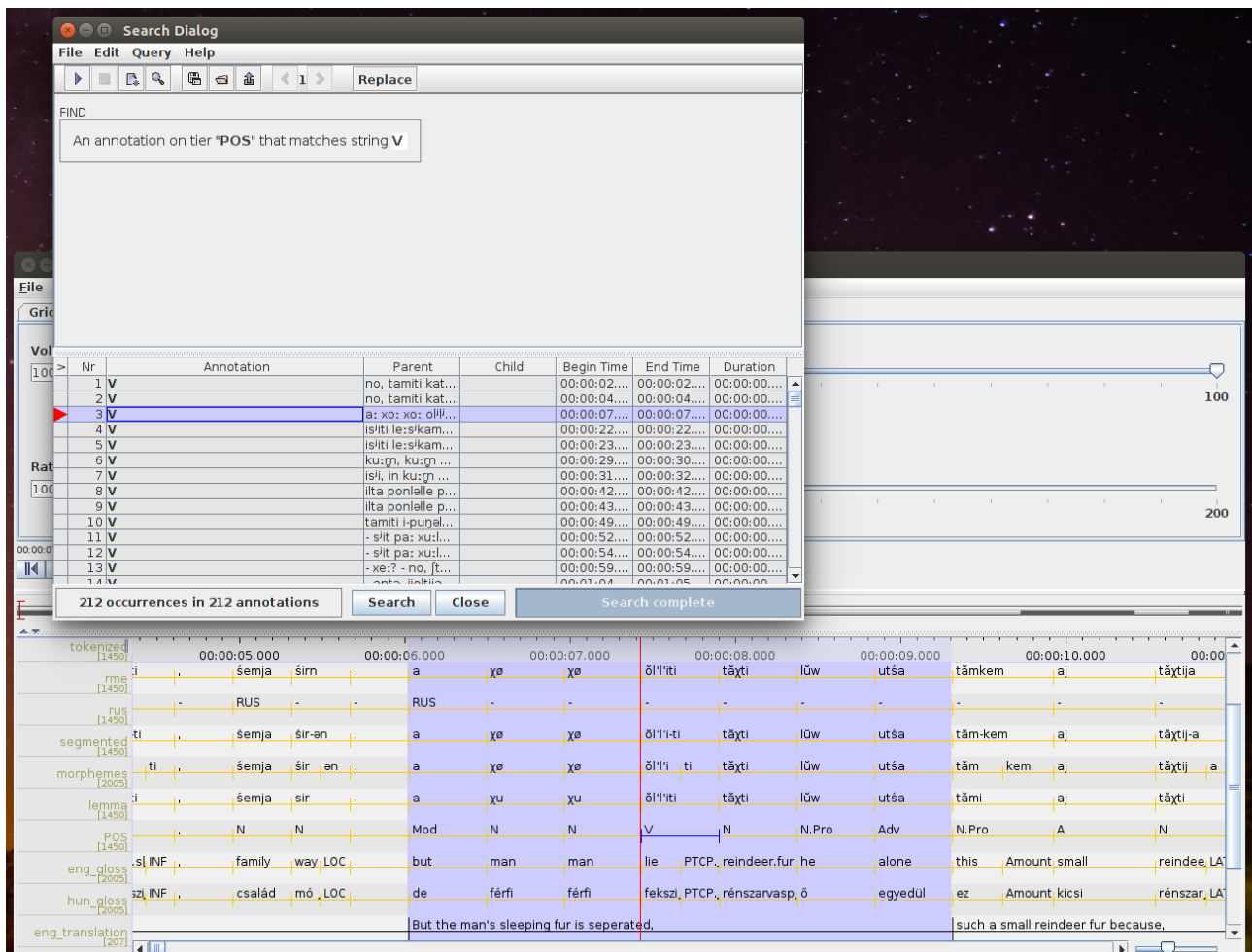
Miután betöltöttük a szövegfájlt az ELAN-ba, egyszerű és komplex kereséseket futtathatunk rajta a felső menüsáv *Search > Find (and Replace)*... menüpontja segítségével, amelyet elérhetünk a *Ctrl+F* billentyűkombinációval is:



A megjelenő *Search Dialog* ablakban megadhatjuk, hogy melyik *tier*-en (tkp. elemzési szinten) történjen a keresés. Az elemzési szintek nyelvenként változnak, például morfológiai elemzés nem érhető el minden uráli nyelv esetében. A *tier*-ek melletti mezőben adjuk meg a keresett szót, kifejezést vagy elemzést. Ha például az adott szöveg igéit szeretnénk listázni, akkor a *tier*-t POS-ra (= part of speech) állítjuk, az elemzést pedig V-re.

A következő sorban opcionálisan megadható az, hogy a fájlban belül hol keressünk perc-másodperc pontossággal. Ennek csak a hanganyaggal rendelkező fájlok esetében lehet jelentősége.

A *Search* gombra kattintva elindítjuk a keresést, amelynek eredménye egy lista a találatok részletes adataival. Ha rákattintunk egy találati sorra, megnézhetjük az eredeti kontextusában, az összes elemzési szinttel együtt:

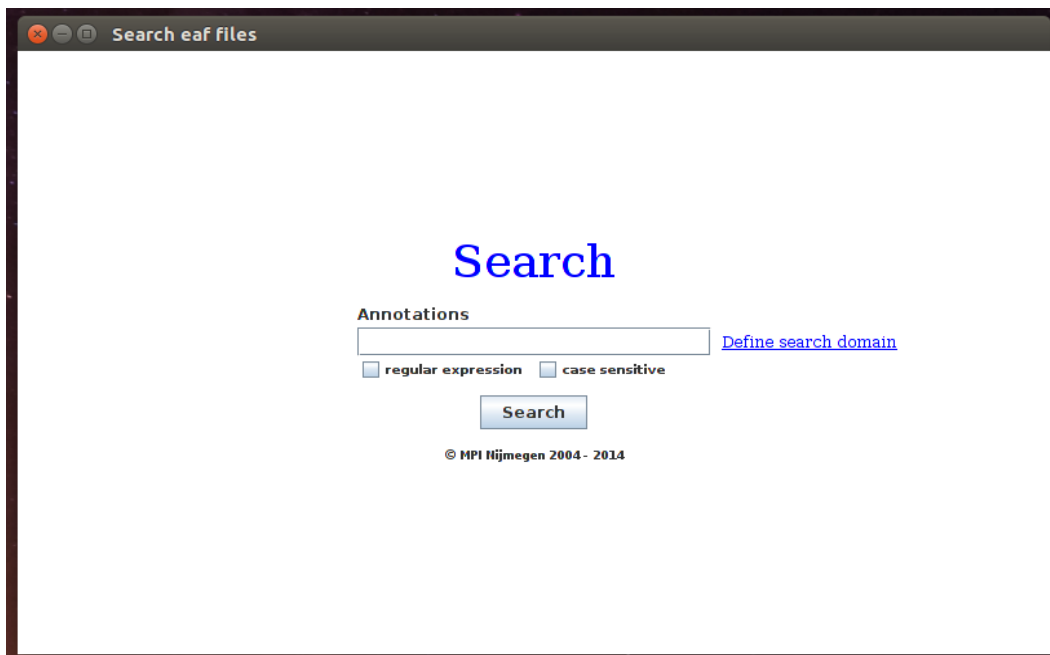


Ezen a felületen lehetőség van ennél komplexebb keresésre is, reguláris kifejezések (regular expressions) és különféle megszorítások (constraints) használatával. Ehhez részletes segítséget nyújt az ELAN használati útmutatójának a keresésre vonatkozó szakasza, amely itt olvasható:

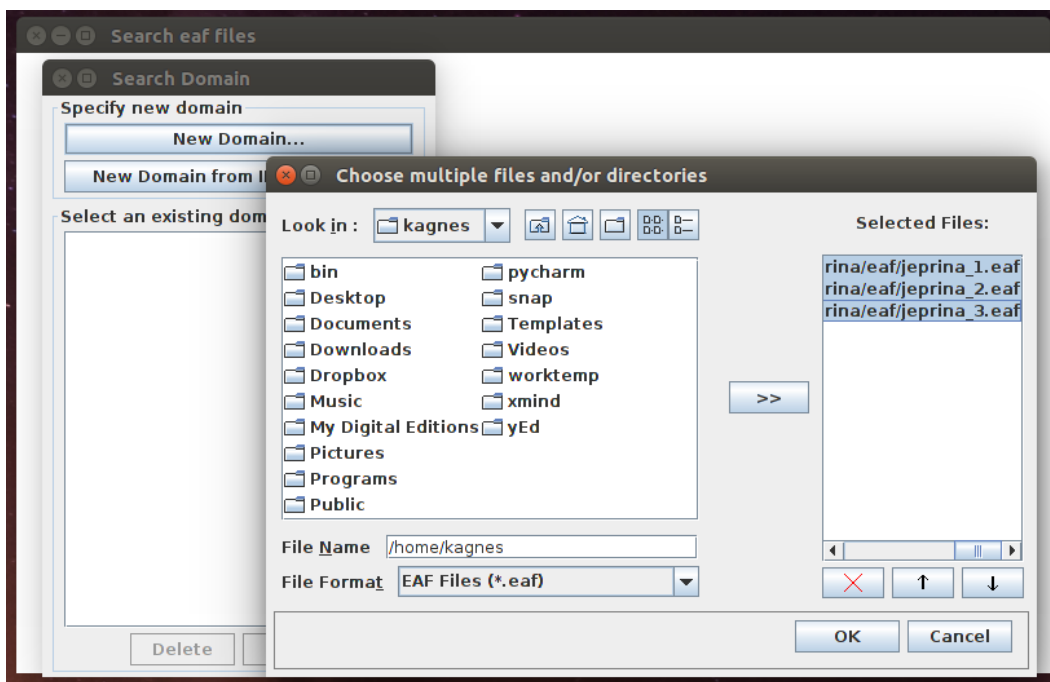
<https://www.mpi.nl/corpus/html/elan/ch04.html>

2.3. Több szövegen belüli keresés

Ha több szövegen belül keresünk, nem kell egyetlen fájlt sem betöltenünk, csak megnyitjuk az ELAN-t. A felső menüsávban a *Search > Search Multiple eaf...* menüpontot választjuk, vagy használhatjuk a *Ctrl+Shift+F* billentyűkombinációt. Ekkor a következő felület nyílik meg:



Kattintsunk a *Define search domain*-ra, majd válasszuk a *New Domain...* opciót (ha első alkalommal futtatunk több szövegen keresést). A felugró ablakban jelöljük ki minden olyan .eaf fájlt, amiben keresni akarunk, majd kattintsunk az *OK* gombra:



Ekkor megjelenik egy ablak, ahol lehetőségünk van nevet adni a választott domainnek és elmenteni, így későbbi használatkor már csak be kell töltenünk a létrehozott domaint.

A domain kiválasztása után az *Annotations* mezőben kell megadnunk a keresett szót, kifejezést vagy elemzést. Itt is van lehetőség komplex keresésre ugyanúgy, ahogy az egy szövegben folytatott keresés esetében. A találati listában meg lehet nézni a találatokat az eredeti kontextusukban. Ez annyiban tér el a korábban bemutatottól, hogy amikor kiválasztjuk a sort, az ELAN kérni fogja annak a fájlnak a betöltését, ami a találati sort tartalmazza (a lenti példában a *jeprina_2.eaf*-ot):

ELAN 5.0.0-beta - jeprina_2.eaf

File Edit Annotation Tier Type Search View Options Window Help

Grid Text Subtitles Lexicon Comments Recognizers Metadat

00:02:52.571 Selection: 00:02:52.571 - 00:02:53.08

Annotations Search eaf files

Annotations Mod.Neg Search Define search domain Export

regular expression case sensitive

1-50 51-53 53 occurrences in 53 annotations in 3 files (0.162 seconds)

Nr	File	Tier	Before	Annotation	After	Parent	Child	Begin Ti...	End...	Dur...
17	jeprina_1	POS		C Mod.Neg	Mod			00:40:2...	00:...	00:...
18	jeprina_1	POS	Adv.Pro	Mod.Neg	V			00:03:1...	00:...	00:...
19	jeprina_1	POS	V	Mod.Neg	Mod			00:36:4...	00:...	00:...
20	jeprina_1	POS	N.kinship	Mod.Neg.Imp	N			00:42:5...	00:...	00:...
21	jeprina_1	POS	Num	Mod.Neg	.			00:39:4...	00:...	00:...
22	jeprina_1	POS	.	Mod.Neg	.			00:01:0...	00:...	00:...
23	jeprina_1	POS	? Mod.Neg	N.Pro				00:01:4...	00:...	00:...
24	jeprina_1	POS	V	Mod.Neg	V			00:04:2...	00:...	00:...
25	jeprina_1	POS	Adv	Mod.Neg	C			00:39:0...	00:...	00:...
26	jeprina_1	POS	.	Mod.Neg.Imp	?			00:43:0...	00:...	00:...
27	jeprina_1	POS	N.Pro	Mod.Neg	N.Pro			00:44:3...	00:...	00:...
28	jeprina_1	POS	A	Mod.Neg	C			00:39:0...	00:...	00:...
29	jeprina_1	POS	###	Mod.Neg	V			00:44:2...	00:...	00:...
30	jeprina_1	POS	V	Mod.Neg	V			00:43:0...	00:...	00:...
31	jeprina_1	POS	Adv-N	Mod.Neg	N			00:42:5...	00:...	00:...
32	jeprina_2	POS	A	Mod.Neg	N			00:01:5...	00:...	00:...
33	jeprina_2	POS	C	Mod.Neg	V			00:02:5...	00:...	00:...
34	jeprina_2	POS	N.kinship	Mod.Neg	N			00:00:4...	00:...	00:...
35	jeprina_2	POS	.	Mod.Neg	.			00:02:5...	00:...	00:...
36	jeprina_2	POS	N	Mod.Neg	V			00:00:2...	00:...	00:...
37	jeprina_2	POS	.	Mod.Neg	.			00:01:2...	00:...	00:...
38	jeprina_2	POS	A	Mod.Neg	Adv			00:03:4...	00:...	00:...
39	jeprina_2	POS	V	Mod.Neg	V			00:01:3...	00:...	00:...

IPA Test

default [65]

tokenized [512]

rme [512]

rus [512]

segmented [512]

morphemes [718]

lemma [512]

POS [512]

eng_gloss [718]

hun_gloss [718]

eng_translation [65]

ger_translation [65]

hun_translation [65]

at sɔ:rl. li

at

ät sorl ? ät

- - -

ät sor-l ? ät

ät sor | ? ät

ät sorti ? ät

Mod.Neg V ? Mod.Neg

no to.dry PRS ? no

nem szárad PRS ? nem

- It doesn't dry out. It is certainly a wet tree. Certainly the exterior

-

- Nem szárad ki. Az hiszen vizes fa. A külseje hiszen