

KOLLÉGIUMI TÁRSALGÁSOK ADATBÁZISSÁ „SZELÍDÍTÉSE”: A BEKK

Szabó Gergely és Vargha Fruzsina Sára
ELTE, UOC NyTI

2019. január 16., Budapest
Hétköznapi interakciók, részvétel, ideológia

AZ ELŐADÁS VÁZLATA

1. Bevezetés – a BEKK
2. A Bihalbocs
3. Annotálás és anonimizálás a gyakorlatban
4. Kitekintés

I. BEVEZETÉS – A BEKK

**Miről
beszélgetnek
a szomszéd
szobában?**

**Hogyan
beszélnek
férfiakról
vagy nőkről?**

**Hogyan
beszélnek
másokról?**

**Hogyan
beszélnek
önmagukról?**

**Budapesti Egyetemi Kollégiumi Korpusz:
2015–2017, hétköznapi társalgások kollégiumokban**
*A beszélők interakcióik során hogyan hoznak létre, tartanak
fenn és alkotnak újra különböző társas kategóriákat és
azokhoz kapcsolódó jelentéseket?*

A BUDAPESTI EGYETEMI KOLLÉGIUMI KORPUSZ

- magyar nyelvű hétköznapi társalgások első nagyobb adatbázisa
- 2015–2017 között készült felvételek
 - az ELTE Tehetséggondozó Tanácsának támogatásával
 - a kutatás résztvevői: oktatók, felsőéves tutorok, hallgatók
 - Magyar Nyelvtörténeti, Szociolingvisztikai, Dialektológiai Tanszék
- Budapesten élő egyetemisták egymással folytatott társalgásai
 - 15 szoba, nagyrészt egynemű csoportok
 - előtanulmány: etnográfiai megfigyelések
 - a társalgások rögzítése után félig strukturált interjú a résztvevőkkel
 - változó minőségű hangfelvételek
 - etikai kérdések: beleegyező nyilatkozatok

A BUDAPESTI EGYETEMI KOLLÉGIUMI KORPUSZ

- korpuszvizsgálat vagy korpusz-támogatott? (vö. Galántai és mtsai 2018)
- társalgáselemzési szempontok előtérben
 - 20 órányi adatbázis
 - a lejegyzés szinkronizálva van a hanganyaggal
 - manuális annotálás és anonimizálás
- az ehhez használt szoftver: BihalBocs



bővebben: <http://www.bekk.elte.hu>, Bodó és mtsai 2017, Bodó 2018

2. A BIHALBOCS

BIHALBOCS ALAPELVEK

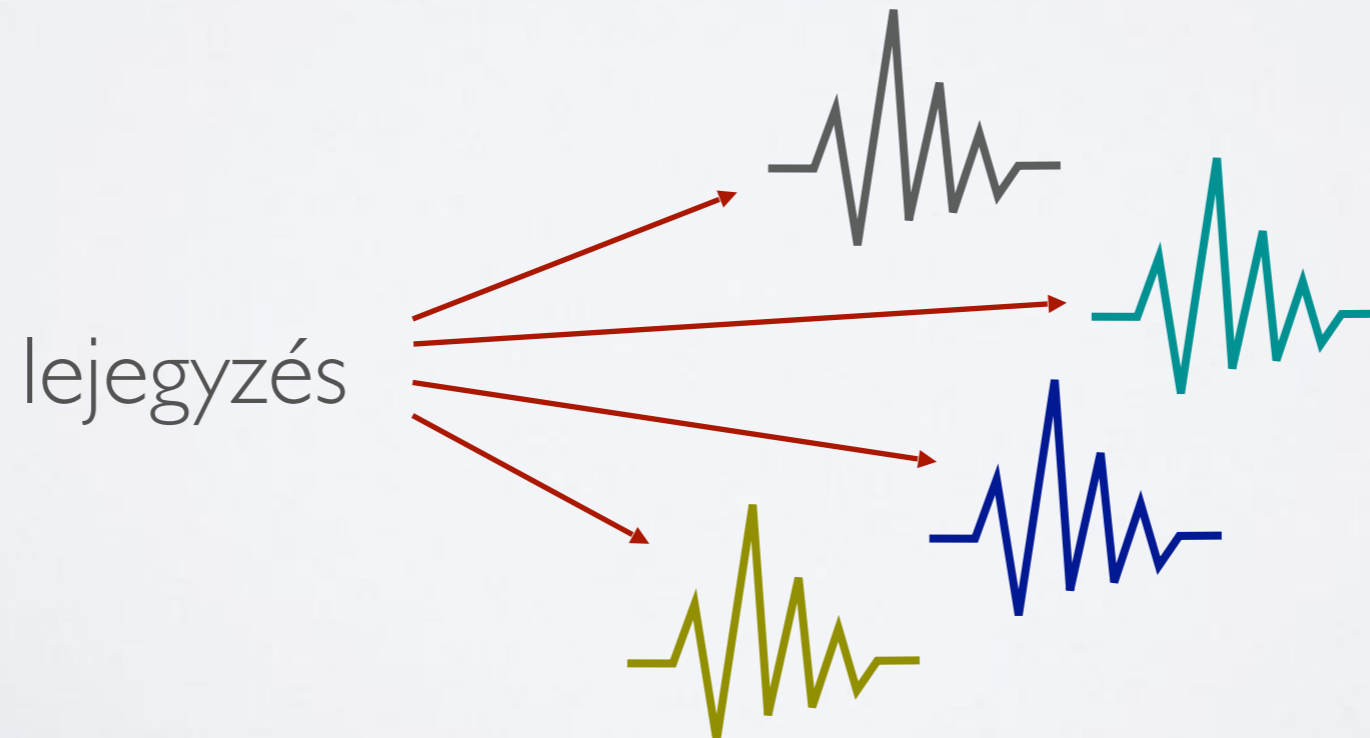
- Lejegyzési standard (magyar egyezményes hangjelölés) támogatása, az informatizált lejegyzés de facto standardja
- Hatékonyság: vizuálisan is támogatjuk a lejegyzési standardot, újrahasznosítható adatokat hozunk létre, hang és lejegyzés összekapcsolás egyszerű és pontos
- Kompatibilitás (exportálás): -> HTML, ELAN, PRAAT, you name it
- Feltételek: speciális informatikai platform, font, illetve forráskód szintű programozás

HANG ÉS LEJEGYZÉS ÖSSZEKAPCSOLÁSA

A lejegyző programok általános logikája:



Hang és lejegyzés viszonya a Bihalboccsban:



LEJEGYZÉS ÉS MARKEREZÉS

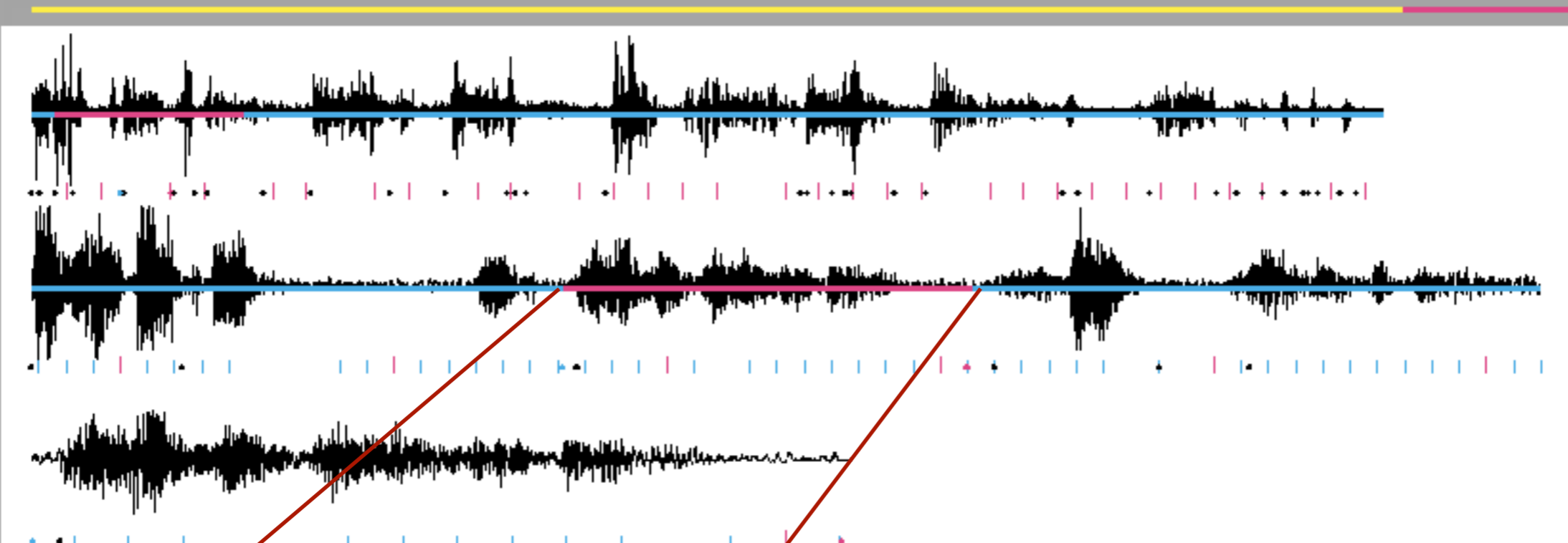
7b Kilép Edit Hang BBText BBAbt. flam_t1_aa.BXT 100% P 16:08

43.06640625 712.44

ment ment... m. nélkül bezár! más.m.txt bxt ? bxt hangf ?

hangrajz + hangot nyit... hangrajzot csuk hang ? 3550. k.tól balra marker: 191 *flam_t1_a.wav* 1 2556284

mki b: 22 22 < kij. vv pötty vtörlés la vtörlés (ennyi visszavont lépést lehet MÉCIS: 0)



```
{1 •<Akkor ne beszélj<ünk semmiről>•?}
{2 •<Azt mon>dták, hogy | hogy•}
{1 •Nem adják ki a rendőrségnek•?}
{3 •(nevet)•}
{2 •De, de, csak a rendőrségnek•.}
{1 •Lehet, hogy besúgó [00] geci vagy amúgy azt csak |• (nevet) •Jelentesz (férfi
utónévl)nak meg i- (n meg a rendőröknek)•.}
```

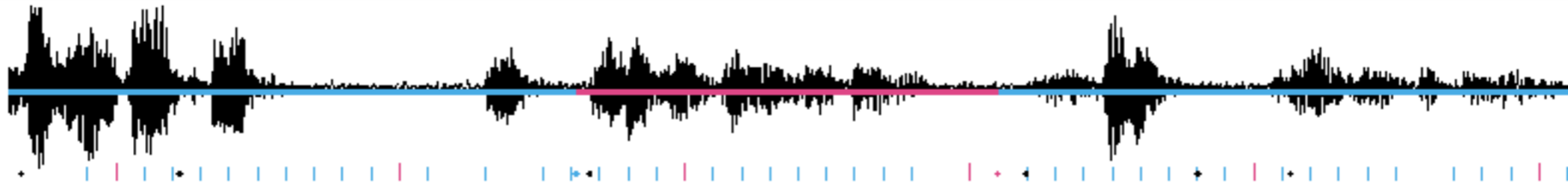
LEJEGYZÉS ÉS MARKEREZÉS

7b Kilép Edit Hang BBText BBAbt. flam_t1_aa.BXT 100% P 18:14

z 43.06640625 ment ment... m. nélkül bezár! más.m.txt bxt ? bxt hangf ?

283 468 hangrajz + hangot nyit... hangrajzot csuk hang ? flam_t1_a.wav 436 288 S: 1 / 5647162 = 4 : 16 : 107

mki b: 22 22 < kij. vv writeSndFile: elapsed secs: 0.013624790008 (ennyi visszavont lépést lehet MÉCIS: 0)



```
{1 •<Akkor ne beszélj<ünk semmiről>•?}
{2 •<Azt mon>dták, hogy | hogy•}
{1 •Nem adják ki| a rendőrségnek•?}
{3 •(nevet)•}
{2 •De, de, csak a rendőrségnek•.}
{1 •Lehet, hogy besúgó [00] geci vagy amúgy azt csak |• (nevet) •Jelentesz (férfi
utónévl)nak meg i- (n meg a rendőröknek)•.}
```

ANONIMIZÁLT RÉSZLET

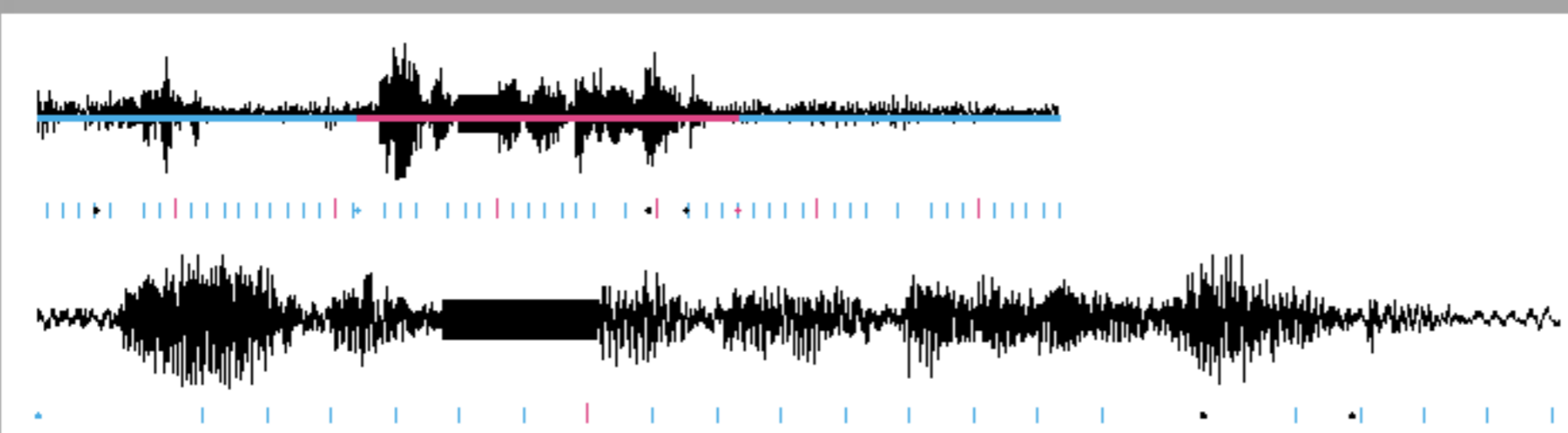
7b Kilép Edit Hang BBText BBAbL. 🔄 🔗 📶 🖥️ 🔊 100% 🇮🇪 P 18:13 🔍 🌐 ☰

flam_t1_aa.BXT

ment ment... m. nélkül bezár! más.m.txt bxt ? bxt hangf ?

hangrajz + hangot nyit... hangrajzot csuk hang ? 1 !! átneve

mki b: 22 22 < kij. vv writeSndFile: elapsed secs: 0.013624790008 (ennyi visszavont lépést lehet MÉCIS: 0)



```
{1 •<Akkor ne beszélj<ünk semmiről>•?}
{2 •<Azt mon>dták, hogy | hogy•}
{1 •Nem adják ki a rendőrségnek•?}
{3 •(nevet)•}
{2 •De, de, csak a rendőrségnek•.}
{1 •Lehet, hogy besúgó [00] gecí vagy amúgy azt csak | • (nevet) • Jelentesz (férfi
utónévl)nak meg i- (n meg a rendőröknek)•.}
```

ANONIMIZÁLÁS

7b Kilép Edit Hang BBText BBAbL.

100% P 16:58

flam_t1_a.BXT

ment ment... m. nélkül bezár! más.m.txt bxt ?

bxt hangf ?

400 257

hangrajz + hangot nyit... hangrajzot csuk hang ?

3821. k.tól balra marker: 207 *flam_t1.wav* 1 2651486

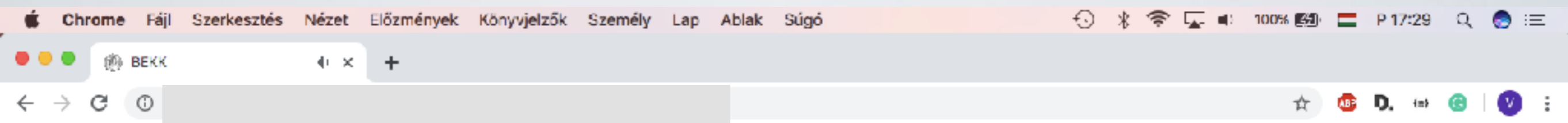
1 !! átneve

mki b: 22 22 < kij. vv ->

vtörítés 1a INYIL KATT. vtörítés 1a KATT.
(ennyi visszavont lépést lehet MÉCIS: 0)

```
{1 [RF] "Ilyen tabutémák vannak amúgy vagy ilyesmi"? "Például ez"?}
{3 "(nevet)"}
{2 "Ja"! "Nem | beszéltünk ilyenek <ről>".}
{1 "Akkor ne beszél<jünk semmiről>"}
{2 "<Azt mon>dták, hogy | hogy"}
{1 "Nem adják ki a rendőrségnek"?}
{3 "(nevet)"}
{2 "De, de, csak a rendőrségnek".}
{1 "Lehet, hogy besúgó [00] geci vagy amúgy azt csak | "(nevet) "Jelentesz
$(férfi utónév1)$nak meg i- (n meg a rendőröknek)".}
{2 "(nevet) "Egyébként arra gondoltam, | hogy le- legtöbbszor úgy indítom majd
el, hogy- hogy ti nem tudtok róla".}
{1 "Jól van, adom amúgy".}
{2 "Jó, mer/"}
{1 "De mér, így is hiteles teljesen".}
{2 "Hát, hiteles, de azért kicsit mégiscsak a diktafonnak szól minden egyes
```


KONVERTÁLÁS HTML-BE



4:16 / 4:52

{1 [Egy \[OO\] köcsögtöcskölő buz.](#) [MR] [\(n Milyen jó szó ez \[OO\] | köcsögtöcskölő\).](#) [/GQ] [/MR]}

{2 [\(nevet\).](#) [Én ma este is színházba megyek, mert én jobb ember vagyok, mint ti.](#) [@.](#)}

{1 [Tudod, hova megyek ma este?](#) [Húzok haza a \[OO\] faszomba.](#)}

{2 [\(nevet\).](#)}

{3 [\(ásít\).](#)}

{1 [RF] [Ilyen tabutémák vannak amúgy vagy ilyesmi?](#) [Például ez?](#)}

{3 [\(nevet\).](#)}

{2 [Ja!](#) [Nem | beszélünk ilyenek «ről».](#)}

{1 [«Akkor ne beszéljünk semmiről»?](#)}

{2 [«Azt mondták, hogy | hogy».](#)}

{1 [Nem adják ki a rendőrségnek?](#)}

{3 [\(nevet\).](#)}

{2 [De, de, csak a rendőrségnek.](#)}

{1 [Lehet, hogy besúgó \[OO\] geci vagy amúgy azt csak \[\\(nevet\\)\]\(#\) \[Jelentesz \\(férfi utónév1\\)nak meg i- \\(n meg a rendőröknek\\).\]\(#\)}](#)}

{2 [\(nevet\).](#) [Egyébként arra gondoltam, | hogy le- legtöbbször úgy indítom majd el, hogy- hogy ti nem tudtok](#)}

KONVERTÁLÁS ELAN-BA

The screenshot displays the ELAN 4.9.4 software interface. At the top, the title bar reads "ELAN_4.9.4_Sierra" and "ELAN 4.9.4 - flam_t1_aa.eaf". The menu bar includes "File", "Edit", "Annotation", "Tier", "Type", "Search", "View", "Options", "Window", and "Help". Below the menu bar, there are tabs for "Grid", "Text", "Subtitles", "Lexicon", "Comments", "Recognizers", "Metadata", and "Controls". The "Text" tab is active, showing a transcription of an audio file. The transcription text is as follows:

ak1
Nyomunk még egy meccset · ? · Egy ucsót · Mm · [OO] «Köcsög gei vagy · Nem [OO] szar, jó · Ja, egy csomó pomót néztem · Nem, gondolom inkognitóba nézte · «Le»galább most már látsz a buzipornón kívül is normális pornót · Mire gondolsz · «A» trap az a legnagyobb műfaj, most kifajtom itt neked · Meg azoknak, akik majd hallani fogják, mer ugye hallani fogják · Azér a legjobb huszonegyedik századi műfaj, mer ötvözi a [DE] hichoppot és a poptot, a két legnépszerűbb műfajt, egyszerre pörgős és érzelmes · · · És | e- hat körülbelül ennyi igazából · Kösz · Kánit? Az it van kinn · Erre gondolsz · Vince, szólaj már meg valamit · · [RF] · Hadd rhögjenek emá, hogy mi van · Hova mesz Viktor, mesélj · Kivel táncolsz · Egyedül · Hát akkor én a helyedbe «csak jövő hét kedden mennék» · Helyedbe csak jövő hét kedden mennék · És egyedül mit csinálsz [OO] bazmeg a péros táncnál · Hanem · De mi, most meg akard [OO] kúmi azt a nőt, vagy mi-miért mentél táncolni · Mm, nem nővel táncolsz · «Meleg vagy, Vince? Vince meleg» · Persze én nem vagyok homofób, nehogy azt higgye valaki · ./ · De semmi baj nincs azzal, hogy meleg vagy, Vince, nekünk elmondhatod nyugodtan · Egy [OO] köcsögtöcskölő buz · · [MR] · (n Milyen jó szó ez [OO] köcsögtöcskölő) · Tudod, hova megyek ma este · ? · Húzok haza a [OO] faszomba · Ilyen tabutémák vannak amúgy vagy ilyesmi · ? · Például ez · «Akkor ne beszéljünk semmiről» | Nem adják ki a rendőrségnek · Lehet, hogy beszélő [OO] gei vagy amúgy azt csak | · (nevet) · Jelentesz (férfi utónév1)nak meg | (n meg a rendőröknek) · Jól van, adom amúgy · De már, így is hiteles teljesen · Aha · Jó | sziasztok ·

Below the text, there is a timeline and a waveform display. The timeline shows a selection from 00:04:16.626 to 00:04:17.103. The waveform display shows the audio signal for "flam_t1_a.wav" with a selected segment highlighted in blue. Below the waveform, there is a transcription grid with three tiers:

Time	ak1 [30]	ak2 [54]	ak3 [30]
00:04:14.000	beszéljünk semmiről»		
00:04:16.000		«Azt mondták, hogy hogy	
00:04:16.000	Nem adják ki a rendőrségnek		
00:04:17.000			(nevet)
00:04:18.000			De, de, csak a rendőrségnek
00:04:19.000			

3. ANNOTÁLÁS ÉS ANONIMIZÁLÁS A GYAKORLATBAN

ANNOTÁLÁS A GYAKORLATBAN

- Etnicitás:
 - EC – cigányokról van szó
 - EM – migránsokról
 - EA - arabokról
 - ED – magyarokról (default)
- Nem:
 - GN – explicit utalás arra, hogy milyenek a nők
 - GF – hogy milyenek a férfiak
 - GQ – nem heteronormatív gyakorlatokról szóló megnyilatkozás
 - GX – a szexualitásról szól

Metanyelv, obszcenitások:

- MR – metanyelvi reflexió
- OM – obszcenitásokra vonatkozó metanyelvi reflexió
- OO – obszcenitás

Grammatikai jelölések:

- DE – értelmezős szerkezet
- DET – értelmezői tagmondat

és egyébek...

ANONIMIZÁLÁS A GYAKORLATBAN

- állandó résztvevők:
pseudonimizálás
- egyéb utónevek: nem és
sorszám
- helynevek

BEKK

Felvétel időpontja: 2015-2017

A felvételt készítette (csoportnév): utóbbiak

Lejegyezte: Wendl Dávid

Annotálta: Kocsis Zsuzsanna

Anonimizálta: Szabó Gergely

1. beszélő: Béla

2. beszélő: Tomi

{1 ●«Úristen»●. ●Meg mi- mindig elmeséli, hogy majd [S] megnéztem (női utónév4)nak a fejezeteit●. [/S]}

{2 ●Ja mer ő most, ja most alapszakos, nem?●}

{1 ●Ja, (dunántúli város)en valami (egyetemi szak neve) szakon, dehát●.}

MIÉRT VAN ERRE SZÜKSÉG?

{3 [OM] ●Főleg úgy, hogy az anyja ma nyakon vágta, mer úgy beszélt vele, mint a kutyával, sőt, még rosszabbul ilyen● ●ez a:- nem az, hogy rondán. [EC] Még a cigányok is néha szebben beszélnek●.}

{1 ●Ilyen tajparszt meg bunkó módon●.}

{3 [S] ●Há mit mondtál, he?● [/S]}

{1 ●Ez a kiáll «a kapanyél a szájából»●.}

4. KITEKINTÉS

KITEKINTÉS

- a nyelvészeti etnográfiai kutatások egyik központi eleme a résztvevők társalgásainak elemzése
- ezek a társalgások viszont sokfelé „burjánzanak”, rizómatikusan szerveződnek
- a „szelídítés” (szinkronizált lejegyzés) azért hasznos, mert a „vad” társalgások mögött megbújó nem-nyelvi jelenségek (szupraszegmentumok, stilizálás) könnyebben hozzáférhetővé válnak a kutatók számára – főleg egy kutatócsoportban

KÖSZÖNJÜK A FIGYELMET!

