



A HuComTech adatbázis

Hunyadi László & Szekrényes István

Debreceni Egyetem

Általános és Alkalmazott Nyelvészeti Tanszék





A multimodális kommunikáció megismerése felé

Korpusztól az adatbázisig

(Hunyadi László)



A HuComTech korpusz 5 éve

A korpusz létrehozásának célja:

- az ember és ember közötti kommunikáció teljesebb megismerése azáltal, hogy kitüntetett figyelmet fordítunk a kommunikáció multimodális természetére



A HuComTech korpusz 5 éve

Multimodalitás a kommunikációban.

- Bár a nyelv talán a legpontosabb kifejezési eszközünk, nem biztos, hogy egyben a leghatékonyabb is:
- Gyakran egy-egy szó, vagy akár egy mondat *szándékolt* értelmét csak akkor kapjuk meg, ha – a kontextus mellett – figyelembe vesszük a nemverbális kísérő elemeket is (hangsúly, prozódia, gesztusok)



A HuComTech korpusz 5 éve

Multimodalitás a kommunikációban

- A nemverbális kifejezési eszközök utalhatnak a beszélő másoktól megkülönböztető, jellemző sajátosságaira, továbbá az adott pillanathoz köthető hangulatára, valamint jellemezhetik a kommunikáció szereplői között kialakult viszonyt, és utalhatnak az esemény szerkezetének a dinamikus alakulására, azaz elősegítik egy esemény teljesebb értelmezését.



A HuComTech korpusz 5 éve

Multimodalitás a kommunikációban

- Bármennyire is adott eseményekhez köthető a nemverbális eszközök *használata*, ezen eszközök és a közöttük lévő viszonyok részét képezik a pragmatikai kompetenciának, így megismerésük hozzájárulhat a kommunikáció formális természetének a megismeréséhez.



A HuComTech korpusz 5 éve

Az így nyerhető ismeretek lehetséges alkalmazási területei:

- nyelvelmélet
- kommunikációelmélet
- pszichológia
- szociológia
- nyelvtechnológia
- információtechnológia
- robotika
- ...



A HuComTech korpusz 5 éve

A korpusz építésében részt vevő modulok:

- kommunikációelmélet
- számítógépes nyelvészet
- nyelvtechnológia
- informatika (digitális képfeldolgozás, adatbázisépítés)
- műszaki tudományok (robotika)



A HuComTech korpusz 5 éve

A korpusz létrehozását irányító intézmény:

- Debreceni Egyetem (BTK, IK)
- szakemberek további intézményekből:
 - MTA Nyelvtudományi Intézete
 - Szegedi Tudományegyetem
 - BME
 - Oului Egyetem
 - King's College London



A HuComTech korpusz 5 éve

A korpusz létrehozását lehetővé tevő források:

- TÁMOP 4.2.2-08/1-20008-0009
- TÁMOP-4.2.2/B-10/1-2010-0024
- CESAR
- TÁMOP-4.2.2.C-11/1/KONV-2012-0002
- TÁMOP 4.2.4.A/2-11-1-2012-0001



A HuComTech korpusz technikai jellemzői

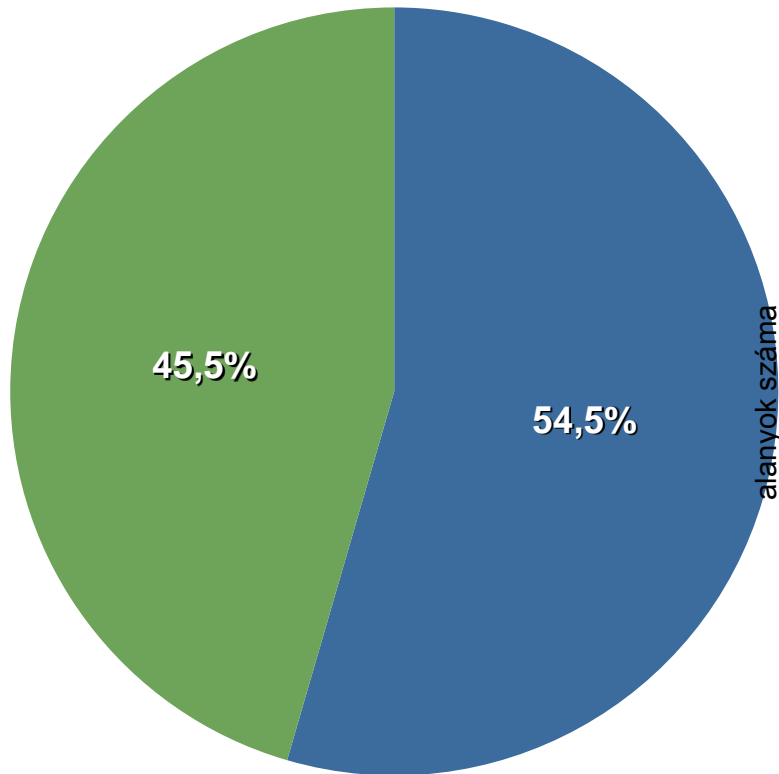
Felvételek

- > 50 óra videofelvétel 3 kamerával és 2 irányított mikrofonnal
- szövegfelolvasás (15 mondat)
- formális dialógus (állásinterjú, ~10 perc)
- informális, irányított dialógus (~15 perc)
- résztvevők száma: 121 (67 férfi, 54 nő, 18-29 év)



A HuComTech korpusz technikai jellemzői

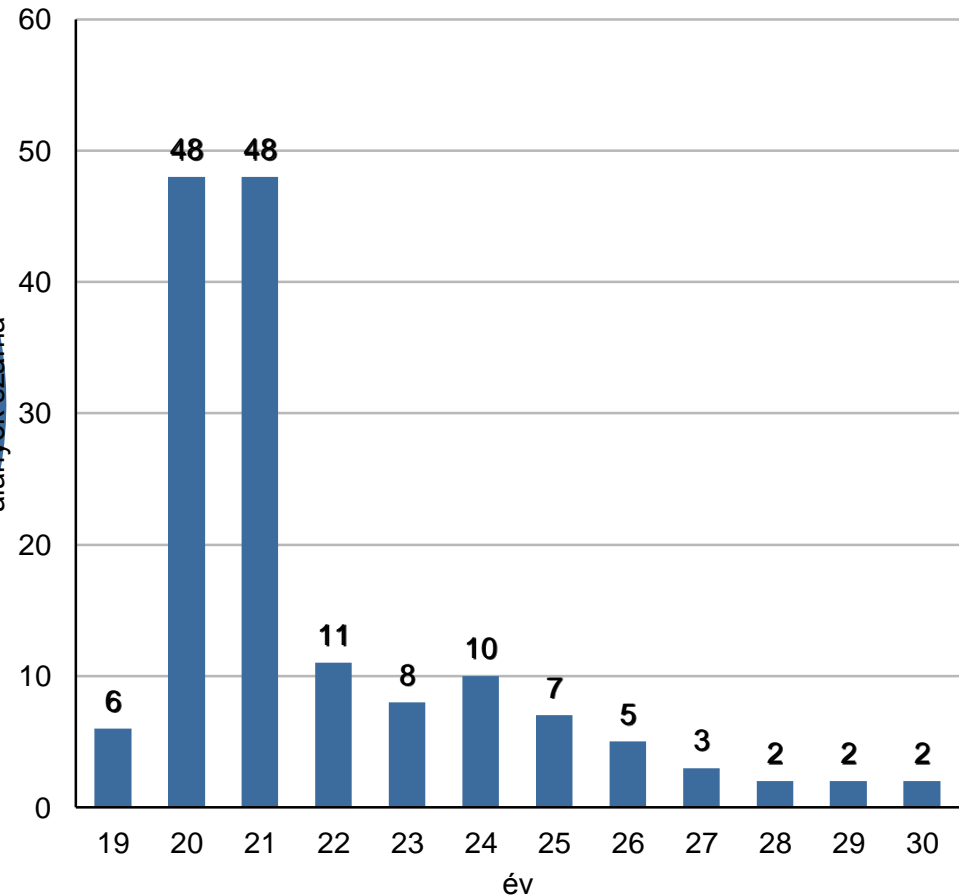
Az alanyok nemek szerinti eloszlása



■ férfi

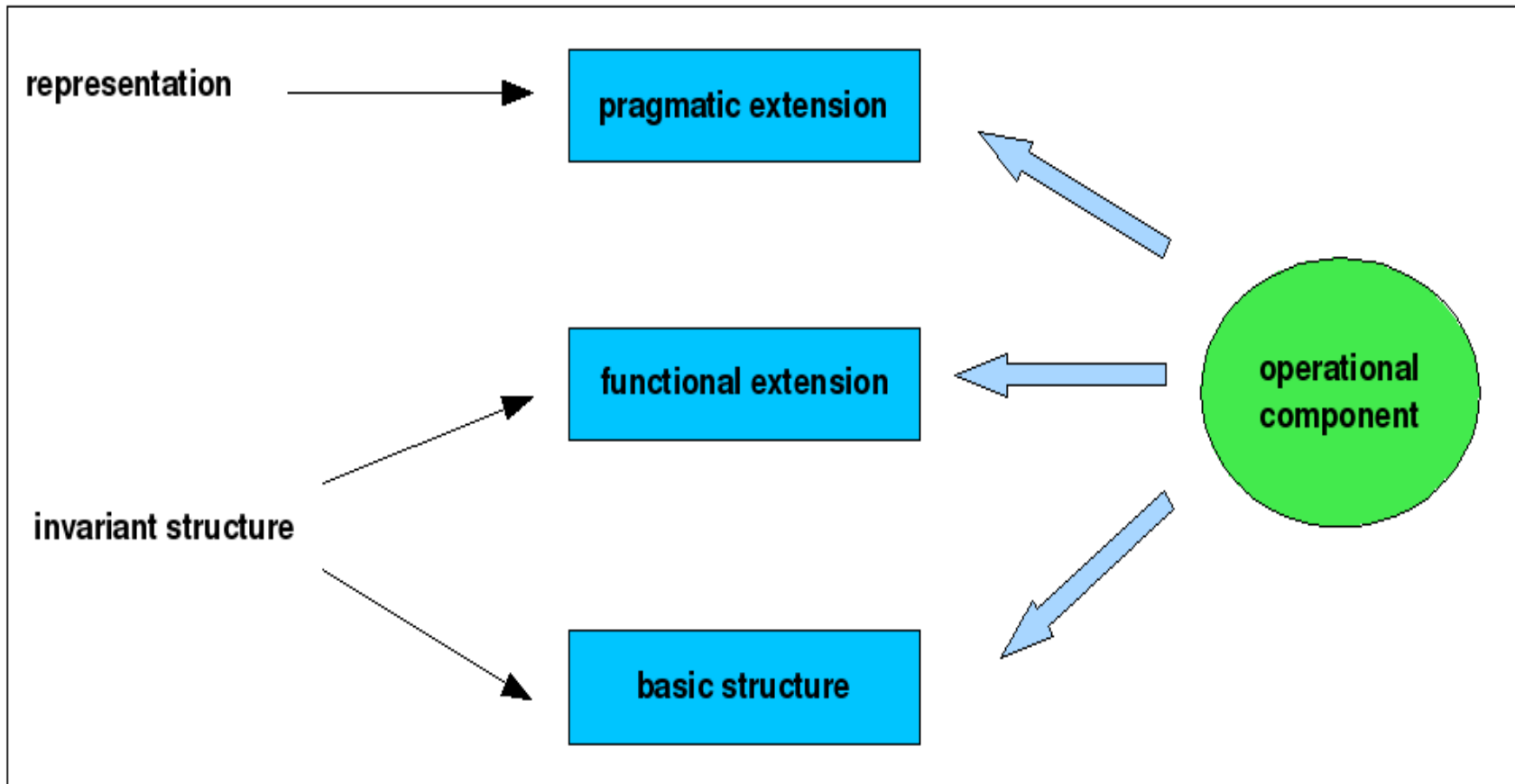
■ nő

Az alanyok életkor szerinti eloszlása



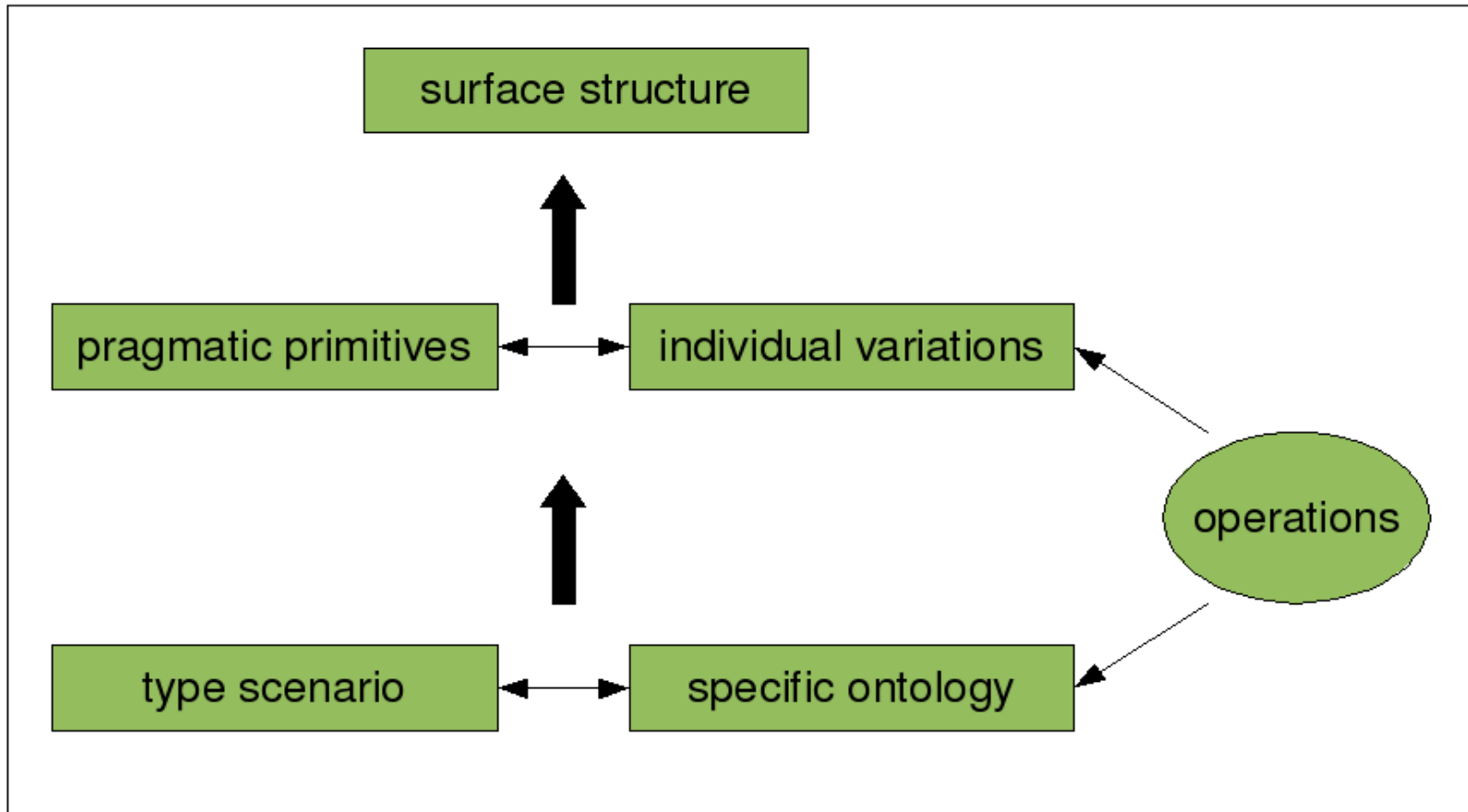
Az annotálást támogató generatív elmélet

- Moduláris



Az annotálást támogató generatív elmélet

- A pragmatikai kiterjesztés



A HuComTech korpusz annotálása

The screenshot displays the HuComTech annotation software interface. At the top, there are three video windows showing different scenes: a man in a red shirt sitting in a chair, a man in a red shirt sitting on a stool, and a woman in a red shirt sitting at a desk. Below the videos is a control bar with playback buttons and a selection mode. The bottom part of the interface is a timeline with multiple tracks for linguistic and non-linguistic annotations.

The timeline tracks include:

- A_discourse** [608]
- A_speaker_text** [608]
- A_agent_text** [485]
- S_syntactic** [614]
- S_sentences** [391]
- P_pitchmovements** [785]
- V_facialExpression** [151]
- V_gazeClass** [463]
- V_eyebrowsClass** [46]
- V_headshiftClass** [562]
- V_handshapeClass** [323]
- V_touchmotionClass** [53]
- V_postureClass** [43]
- V_deicticClass** [3]

The timeline shows a selection from 00:01:26.982 to 00:01:30.674. The selected segment contains the following annotations:

- A_speaker_text**: j%a, <ja>, értem.
- A_agent_text**: {b} %m hát még %o amiatt jöttem Debrecenbe.
- S_syntactic**: 1. 1.0.0.0.0.14 1.0.0.0.0 1.0.0.0.0.0
- S_sentences**: s33 s34 s35 s36
- P_pitchmovements**: rise fall rise ris
- V_facialExpression**: happy+reduced natural+moderate
- V_gazeClass**: right|blink|forwards
- V_headshiftClass**: raise|turn+|lower-turn+|nod
- V_handshapeClass**: both-crossing-fingers|both-other
- V_postureClass**: lean-back|upright

Az audió annotálása

- Elemi annotálás (szünettől szünetig) – jellemzi az annotálás “evolúcióját”
 - a szöveg átírása
 - megakadások, újraindítás, együttes beszéd, idegen szavak, nem érthető szöveg stb.
 - mondattan: főmondat, mellékmondat, alárendelés, beágyazás, beékelés
 - diskurzus: fordulók (turn management)
 - értelmezett érzelmek
 - nem tartalmaz szószintű időjelölést (timestamp)



A videó annotálása

- arckifejezések (érzelmek; az audió nem hallható)
- tekintet
- szemöldök
- fej
- kéz
- érintés
- testtartás
- deixis
- érzelmek (az audió hallható)
- figyelem, egyetértés (audió nélkül)



Pragmatikai annotáció: multimodális

- kommunikatív aktus (konstatív, direktív, kommisszív, elismerés, indirekt)
- támogató aktus (visszacsatolás, udvariasság, javítás)
- tematikus kontroll (topik-kezelés: kezdés, kifejtés, váltás)
- információ (új információ)



Pragmatikai annotáció: unimodális

- fordulóváltás (beszélni kezd, szándékozik beszélni kezdeni, beszélés befejezése)
- figyelem (odafigyelés, figyelemfelhívás)
- véleményegyezés (árnyalatai)
- információ (új információ)
- deixis



Szintaktikai annotálás

- a beszéd szintaktikai szerkezete (adaptált és alulspecifikált leíró szempontok)
- alapelemek: tagmondat, mondat
- a tagmondatok hierarchikus elrendezése
- külön figyelem a hiányos szerkezetekre
- valamely egy kötelező argumentum hiánya
- több, mint egy kötelező argumentum hiánya



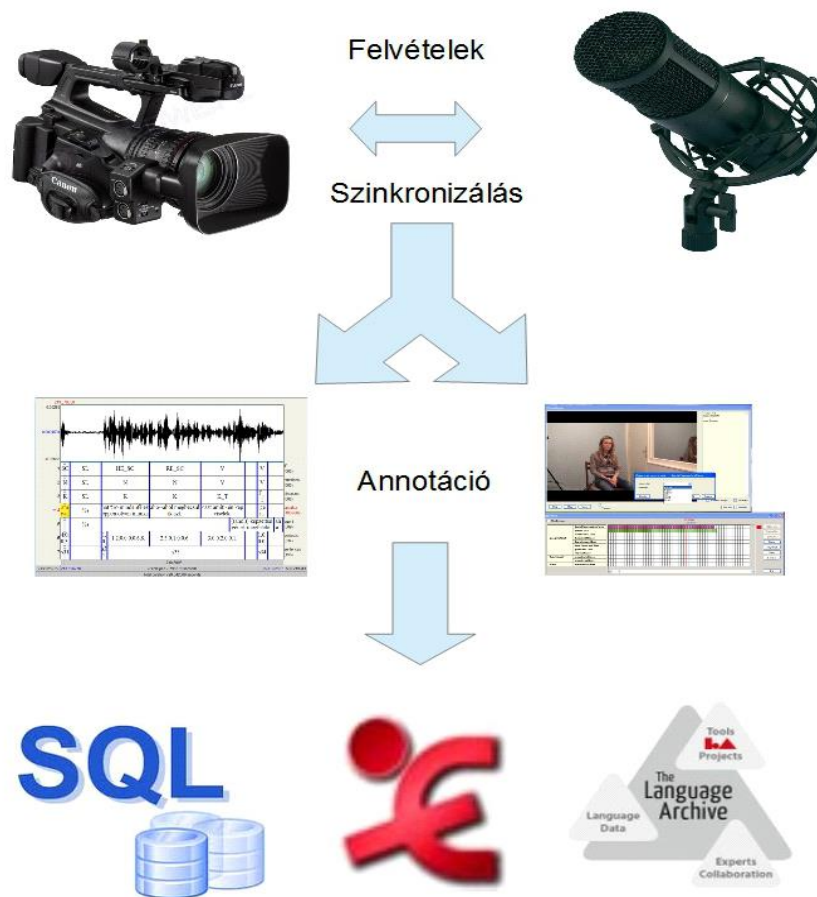
A korpusz feldolgozása

Manuális és automatikus annotációs módszerek

(Szekrényes István)



A korpusz feldolgozása

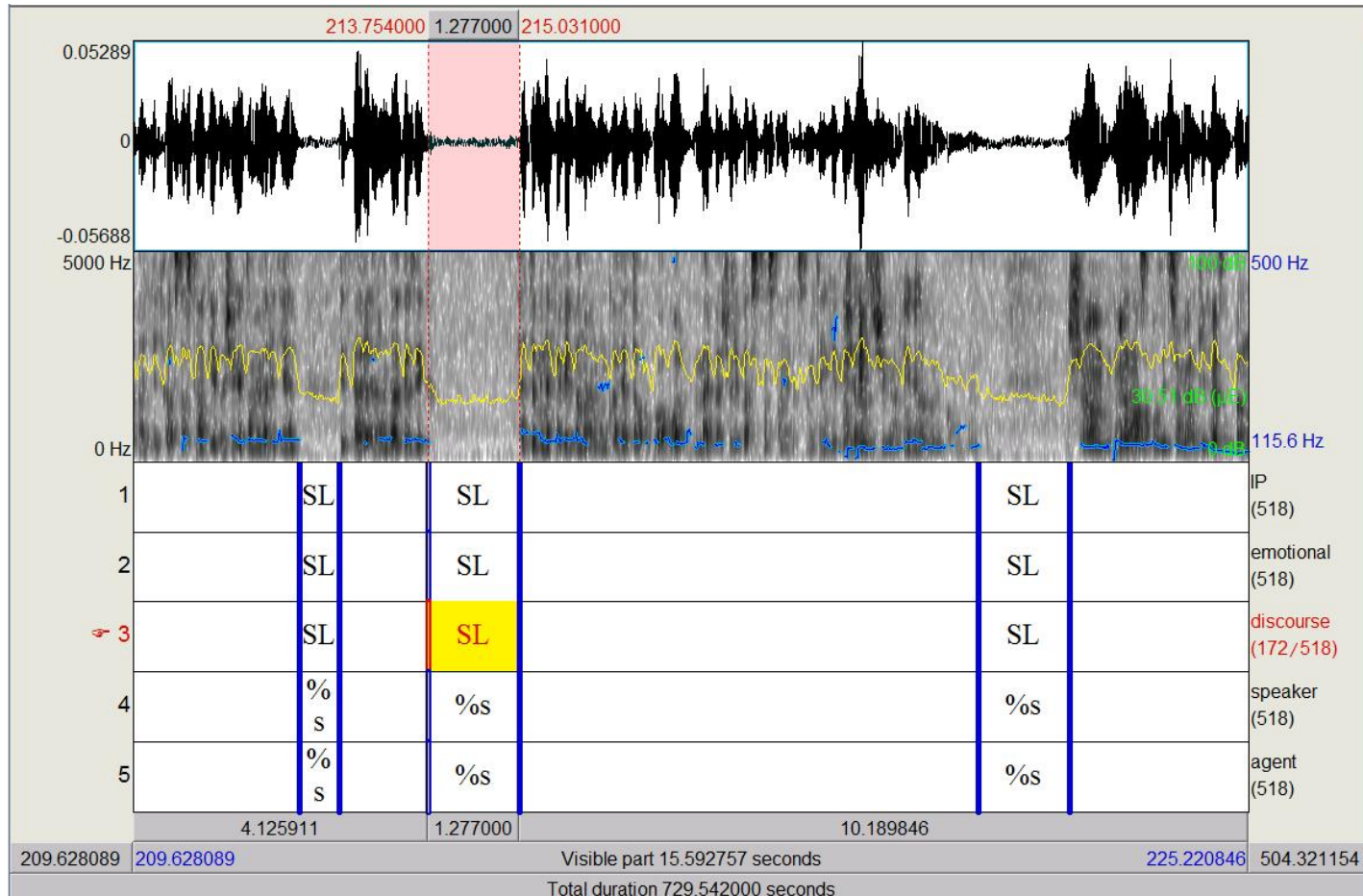


A hanganyag annotációja

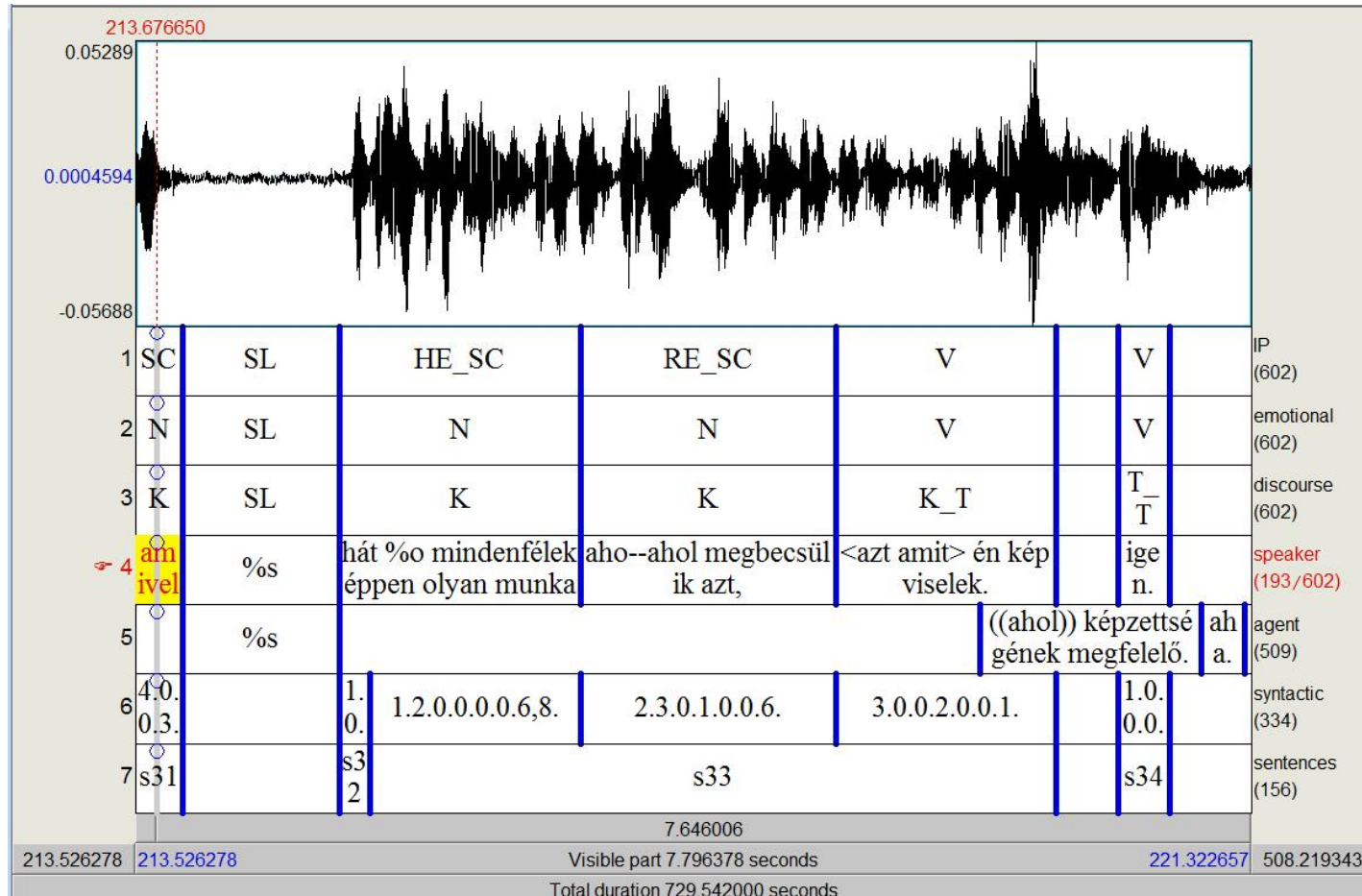
- Szünetek automatikus detektálása (Praat)
- Manuális annotáció (Praat)
- Automatikus és manuális ellenőrzés (Praat)



Szünetek detektálása



Annotációs felület (Praat)



Az annotációk ellenőrzése

- Automatikus :
 - A szimbólumok használatának ellenőrzése
 - A szegmentáció egységességének ellenőrzése
 - Szintaktikai ellenőrzés:
 - indokolatlan ismétlődések
 - az annotációs szintek közötti ellentmondások kiszűrése
 - Anonimizálás
- Manuális:
 - Az automatikusan detektált hibák javítása
 - Minden más...



Szintaktikai annotáció

Feldolgozás előtt:

1.0.0.0.0.14. 1.2.0.0.0.4,6. 2.0.0.1.3.1. 3.0.0.0.2.14,15. 2.0.0.1.3.1.

Feldolgozás után:

1.0.0.0.0.0.14. 1.2_a,2_b.0.0.0.0.4,6. 2_a.0.0.1.3.0.1. 3.0.0.0.2_a,2_b.14,15. 2_b.0.0.1.3.0.1.

Elvégzett műveletek:

- A beékelések, beágyazások, alá- és mellérendelések ellenőrzése
- A beékelések és beágyazások által lineárisan szétdarabolt tagmondatok szintaktikai hierarchiába való rendezése (indexelés által)

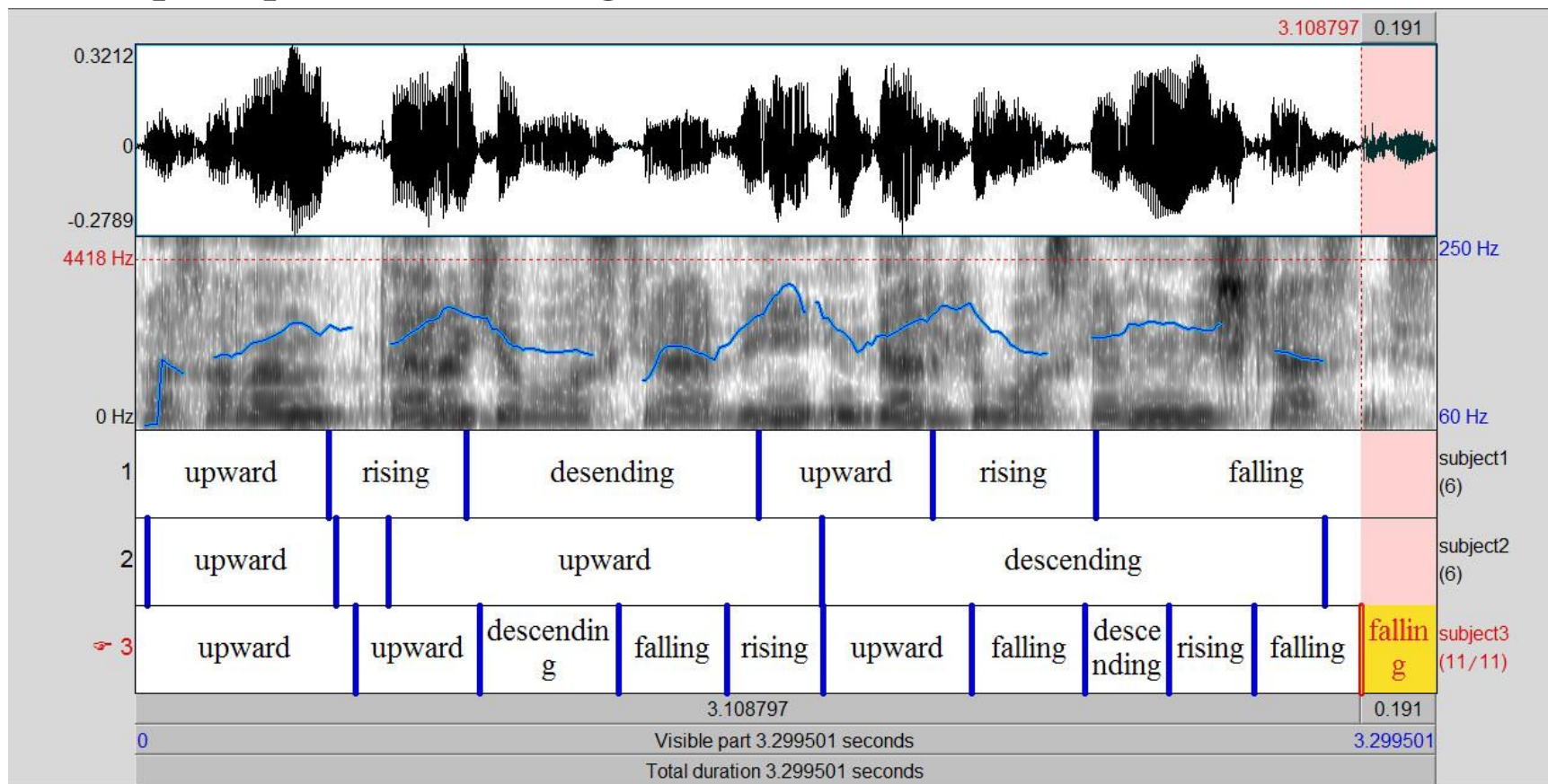
Automatikus prozódia annotáció

- Beszéddallam
- Intenzitás
- Beszédtempó

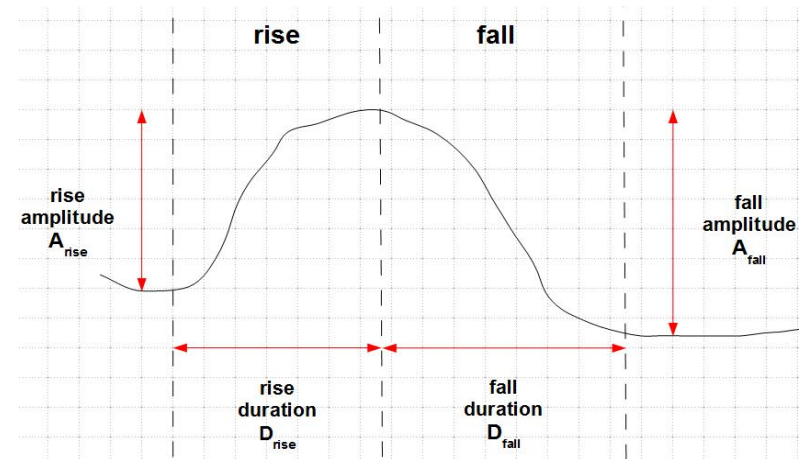
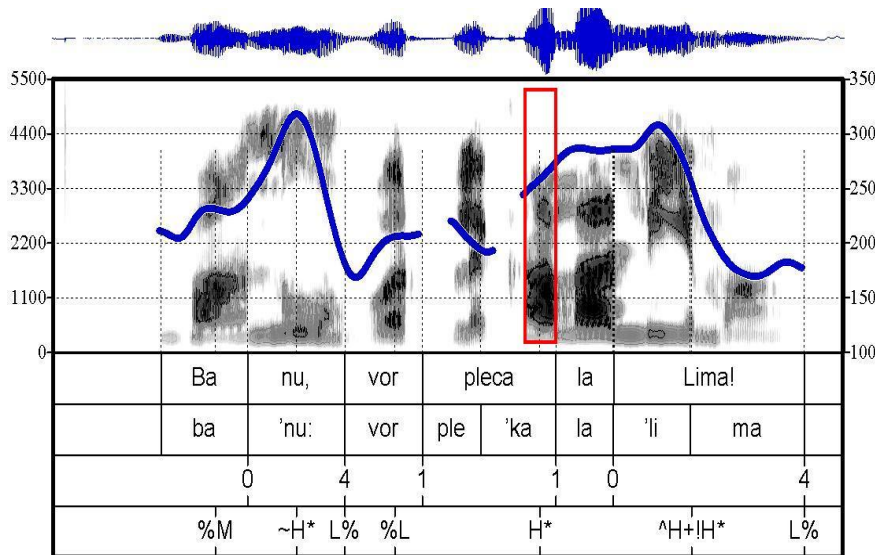
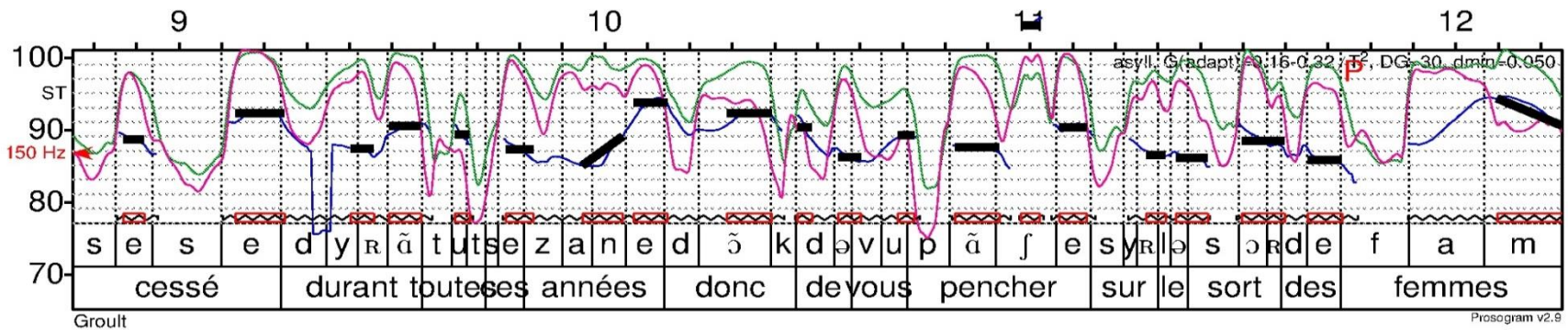


A beszéddallam annotációja

Cél: a percepciónak való megfeleltetés

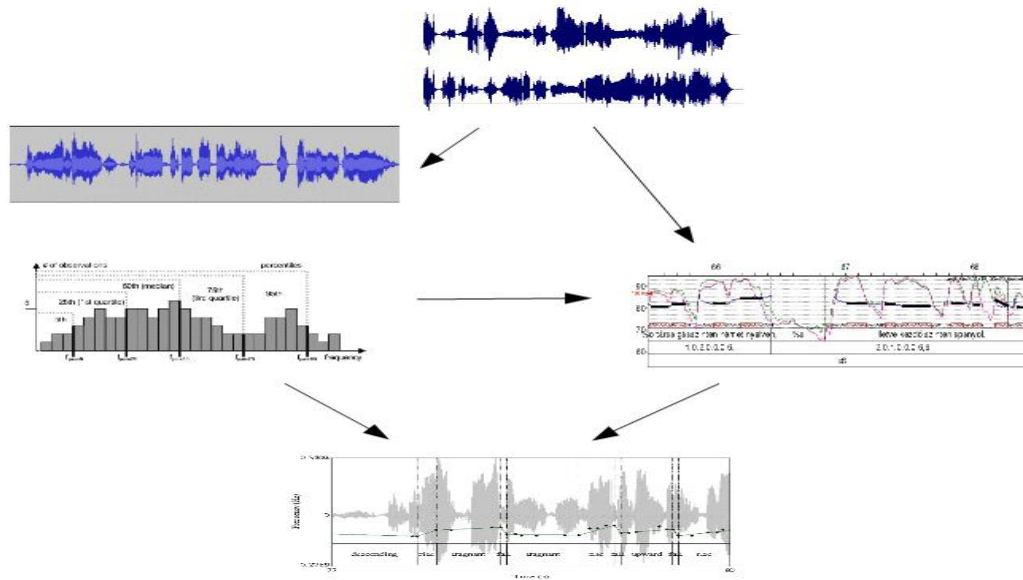


A beszéddallam annotációja

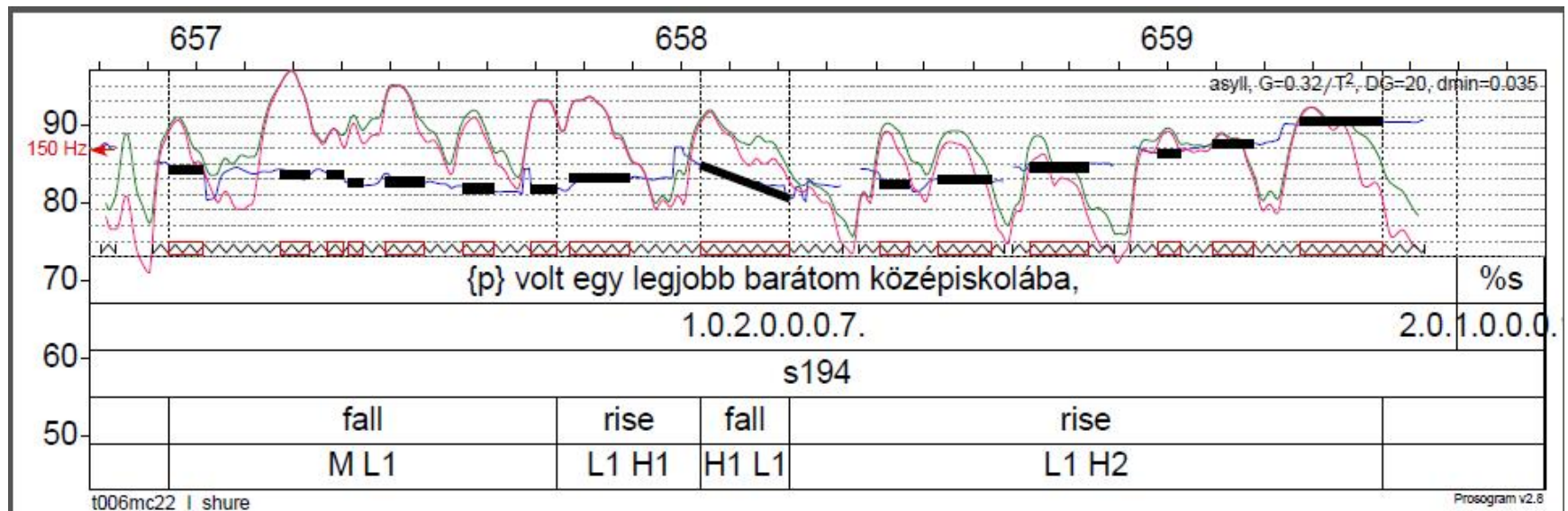


A beszéddallam annotációja

1. A beszélők szétválasztása (az annotációk alapján)
2. Tartományhatárok és küszöbértékek számítása (globális lekérdezések)
3. Az alapprofrekvencia stilizálása (Prosogram)
4. A szótagszintű stilizációk integrálása nagyobb egységekbe (Prosotool)
5. A dallammozgások osztályozása és címkézése (Prosotool)



A beszéddallam annotációja



A videóanyag annotációja

- Manuális annotáció (Qannot)
- Érzelmi kifejeződések automatikus annotációja (Noldus FaceReader)
- Manuális ellenőrzés



Annotációs felület (Qannot)

The screenshot displays the Qannot software interface, which is used for video annotation. It consists of two main windows:

- VideoPlayer:** Shows a video frame of a person sitting in a chair. A checkerboard mask is applied to the person's face. A dialog box titled "Create or delete label - facialExperssionClass" is open, allowing the user to select an expression from a list: natural, happy, suprise, sad, recall, tense. The dialog also includes fields for "expression" and "intensity", and buttons for "Delete", "Ok", and "Cancel".
- QANNOT:** Shows a timeline view of the video. The timeline is divided into columns representing different annotation classes. The classes are organized into three categories:
 - pragmatic:** facialExperssionClass, gazeClass, eyebrowsClass, headshiftClass, handshapeClass, touchmotionClass, postureClass, deicticClass
 - functional:** emotionsClass, emblemsClass
 - basic:** comeventClassThe timeline shows colored bars representing annotations for these classes. The current time is 00:30:00, and the video duration is 00:30:35. The interface includes playback controls (Stop, Play, Pause), volume, and buttons for "Set start", "Set end", "Open Qnt", "Open Video", "Set rules", "Reset", "Play Part", "Save", "Save as", and "Exit".

Szinkronizálás (ELAN)

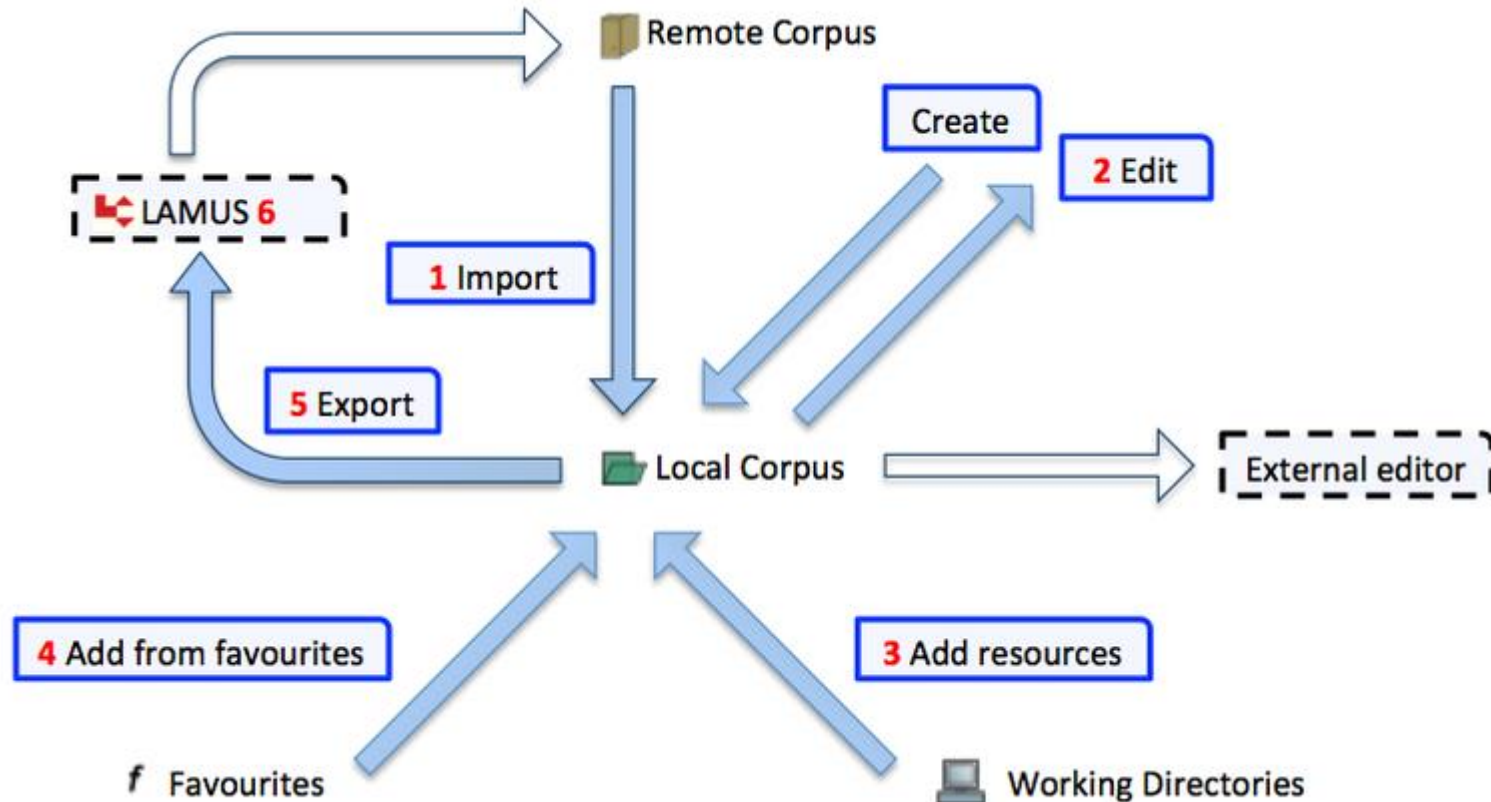
The screenshot displays the ELAN software interface for audio synchronization. At the top, a menu bar includes File, Edit, Annotation, Tier, Type, Search, View, Options, Window, and Help. Below the menu is a timeline from 00:00:00.000 to 00:00:06.000. The main area is divided into three video windows:

- Player 1:** Shows an audio waveform for file 063fv20_F_shure.wav. Offset: 00:00:00.000.
- Player 2:** Shows a video frame of a chair in a room for file 063fv20_F_C2.mp4. Offset: 00:00:00.000.
- Player 3:** Shows a video frame of a person at a laptop for file 063fv20_F_C3.mp4. Offset: 00:00:07.811.

At the bottom, there is a control panel with the following sections:

- Offset:** Radio buttons for "Use Absolute Offsets" (selected) and "Use Relative Offsets". Buttons for "Apply Current Offsets" and "Reset all Offsets to 0".
- Player:** Radio buttons for "All" (selected), "Player 1", "Player 2", and "Player 3".
- Timeline:** Playback controls (stop, play/pause, previous, next, first, last) and a selection range of 00:00:00.000 - 00:00:00.000 0.
- Volume:** A slider set to 100.
- Rate:** A slider set to 100.

The Language Archive



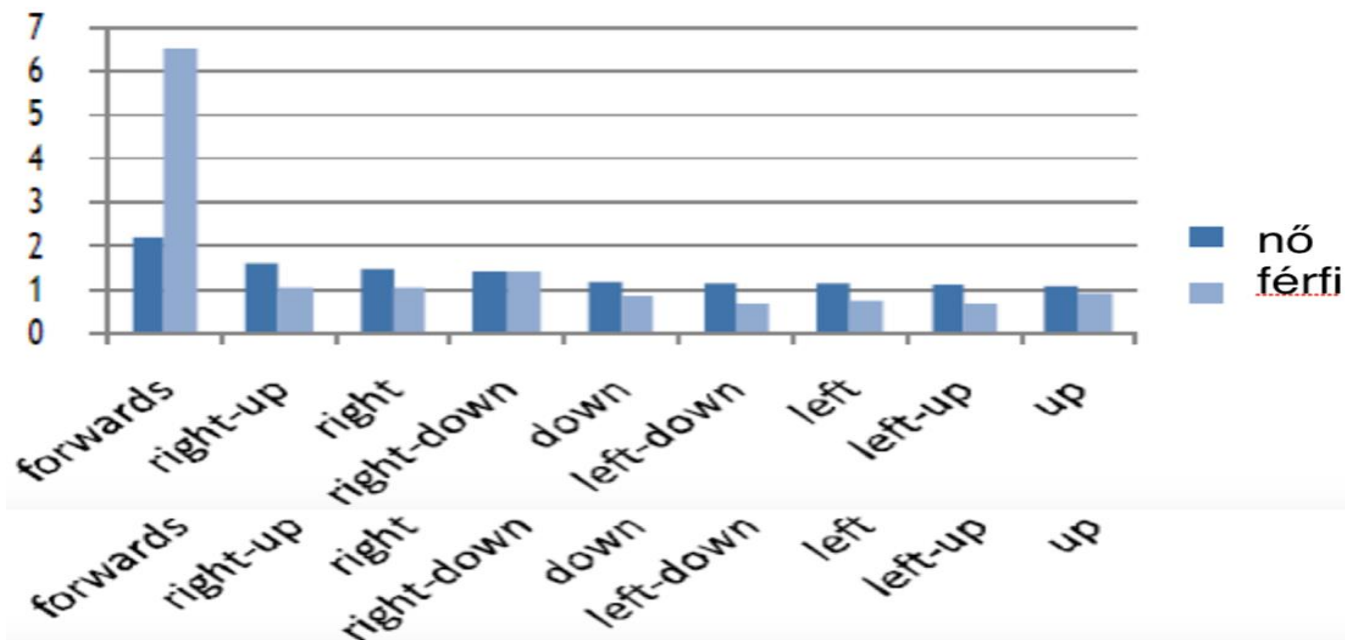


Adatok a HuComTech korpusból



Adatok a HuComTech korpuszból

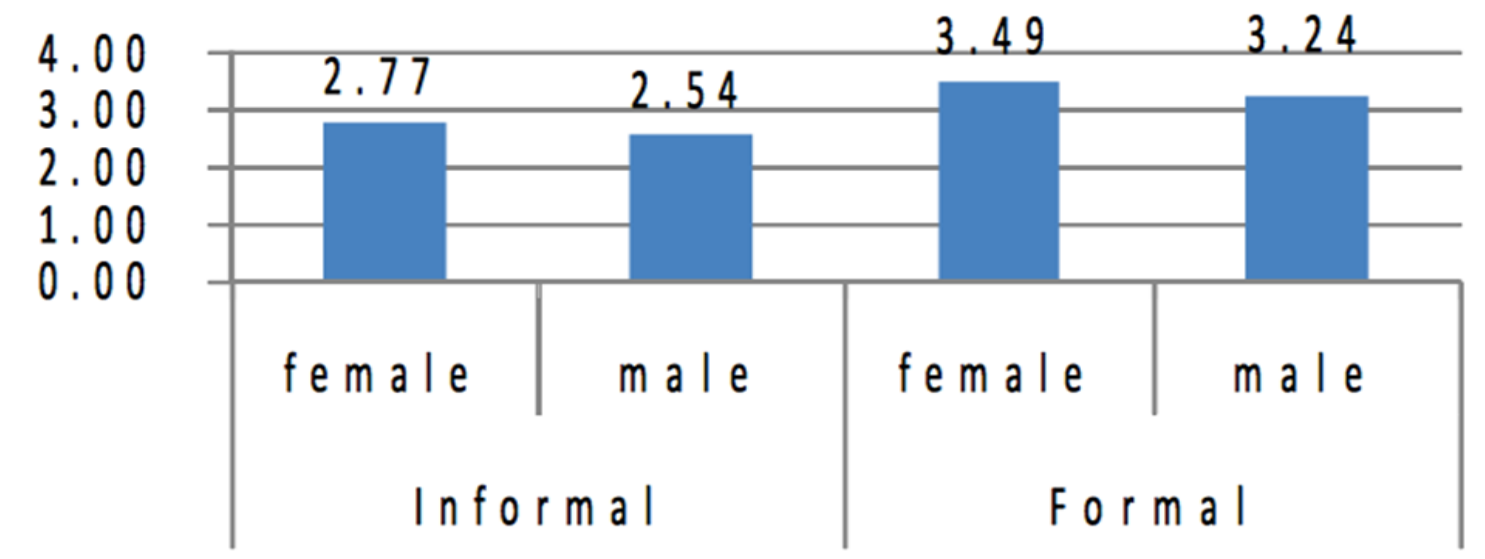
Tekintetirányokat jelző címkék átlagos időtartama forduló átadása során



(Abuczki Ágnes)

Adatok a HuComTech korpuszból

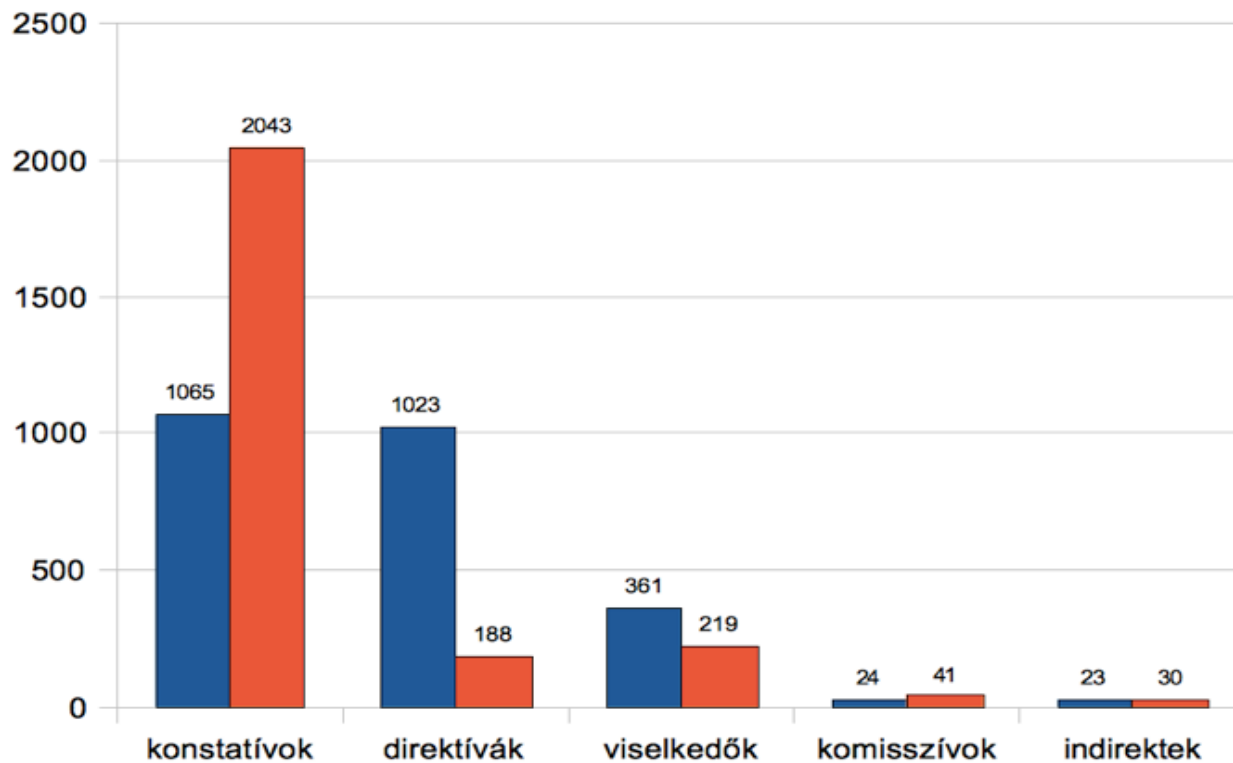
Folyamatos beszéd időtartam egy fordulón belül



(Abuczki Ágnes)

Adatok a HuComTech korpuszból

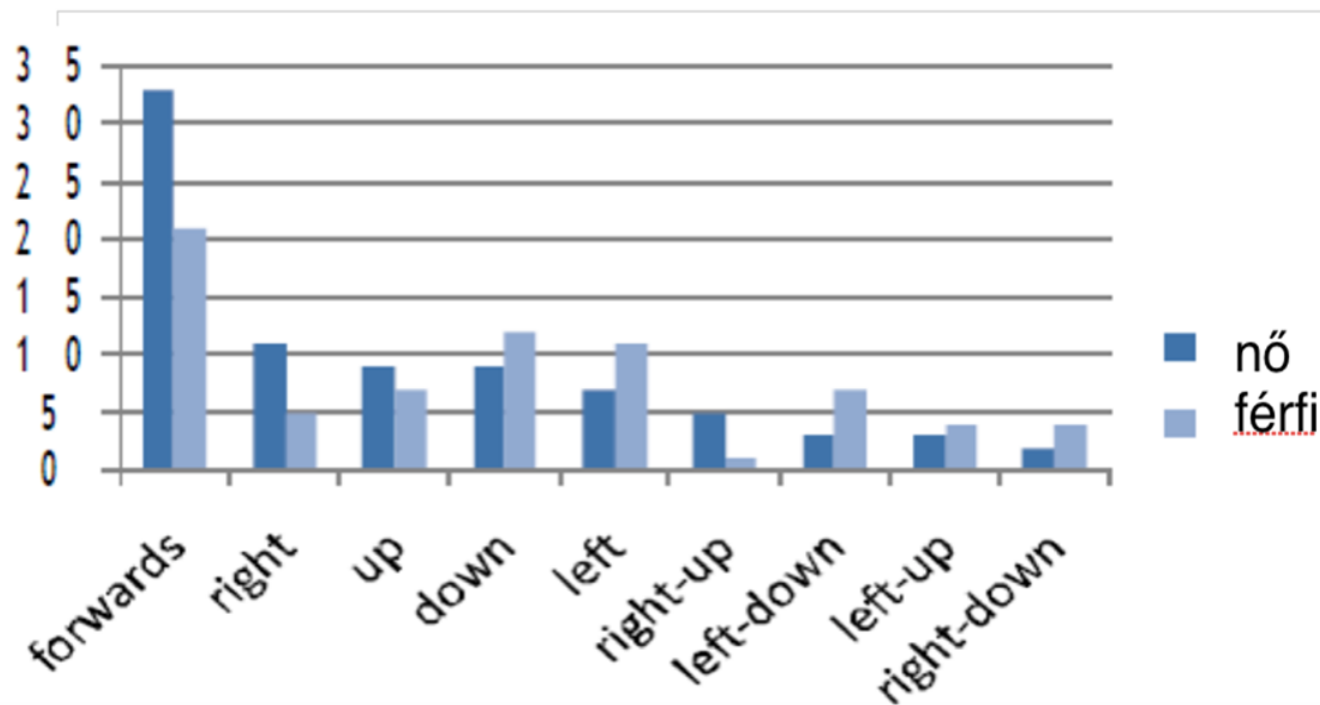
Kommunikatív aktustípusok - multimodális pragmatika



(Abuczki Ágnes)

Adatok a HuComTech korpuszból

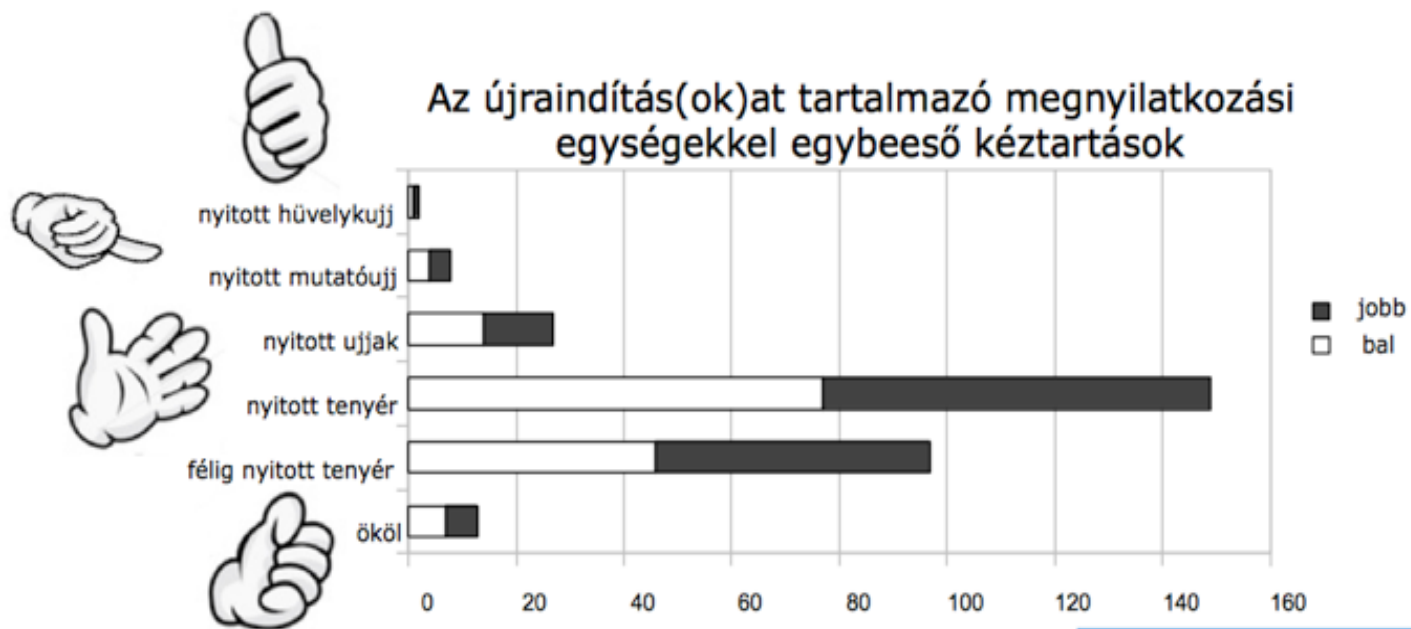
Forduló átadása közben előforduló tekintetiránymintázatok



(Abuczki Ágnes)

Adatok a HuComTech korpusból

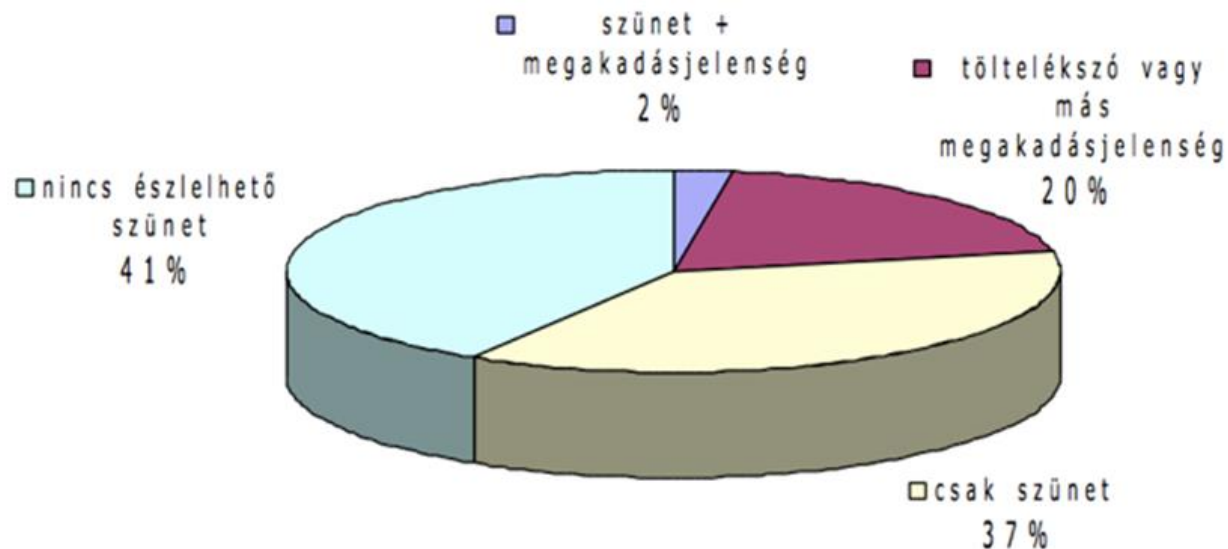
Újraindításokkal egybeeső kéztartások



(Staudt Alexandra)

Adatok a HuComTech korpusból

Újraindított szót megelőző jelenségek



(Staudt Alexandra)

Adatok a HuComTech korpusból

Gyakori beszéddallam-minták

- rise + fall (2904)

	rise	fall	rise	fall
average duration	291.58	403.34	336.35	320.26
standard deviation	327.59	459.35	371.23	404.24

- rise + rise (2)

	rise	rise	rise	rise
average duration	160.5	486.5	360.5	327.5
standard deviation	4.5	146.5	279.5	42.5

Adatok a HuComTech korpusból

Gyakori beszéddallam-minták

- fall + fall (5)

	fall	fall	fall	fall
average duration	184.25	321.75	194.5	435.5
Standard deviation	75.13	145.66	76.63	364.24

- fall + rise (2327)

	fall	rise	fall	rise
average duration	397.58	336.73	389.88	337.94
Standard deviation	464.01	368.27	470.21	381.64

Adatok a HuComTech korpuszból

Ismétlődő beszéddallam-minták

- fall + level + rise (328)

	fall	stagnant	rise	fall	stagnant	rise
average duration	179.36	545.82	205.79	188.93	580.35	173.81
Standard deviation	195.35	704.17	218.54	174.13	698.99	202.61

- rise + level + fall (78)

	rise	rise	fall	rise	rise	fall
average duration	160.19	726.48	194.53	177.83	674.23	129.62
Standard deviation	136.63	872.52	210.12	173.82	733.54	92.18

Adatok a HuComTech korpusból

Ismétlődő beszéddallam-minták

- rise + rise + fall (42)

	rise	rise	fall	rise	rise	fall
average duration	282.05	313.39	358.76	405.71	183.22	238.12
Standard deviation	279.06	297.46	399.16	363.03	182.60	270.15

- fall + fall + rise (46)

	fall	fall	rise	fall	fall	rise
average duration	177.13	564.58	237.51	314.96	422.47	288.29
Standard deviation	22.44	535.97	219.60	399.60	460.60	374.34

Adatok a HuComTech korpuszból

Ismétlődő beszéddallam-minták

- fall + level + rise + level (242)

	fall	stagn.	rise	stagn.	fall	stagn.	rise	stagn.
average duration	118.91	215.41	143.21	359.71	116.83	217.75	140.20	425.22
Standard deviation	80.19	249.27	120.52	455.14	78.53	300.83	131.27	526.93

- rise + level + fall + love (171)

	rise	stagn.	fall	stagn.	rise	stagn.	fall	stagn.
average duration	152.79	410.51	112.15	231.62	160.17	319.48	117.88	254.9-
Standard deviation	146.42	482.48	81.02	330.51	254.04	422.58	78.01	349.87



Adatok a HuComTech korpuszból

- Érzelem + beszéddallam
- Arckifejezést F0-mozgás követ (< 500 ms)

emotion + pitch	fall #	fall (%)	rise #	rise (%)	stagn. #	stagn. (%)	descend. #	descend. (%)	upward #	upward (%)
natural	494	32.76	497	32.26	359	22.82	84	5.26	74	4.62
happy	540	35.98	504	32.79	282	17.96	85	5.35	90	5.65
sad	6	42.86	4	7.03	0	0	3	4.7	1	1.46
surprised	46	20.09	138	55.4	30	9.85	5	1.59	10	3.17
recall	453	34.32	408	30.13	341	24.63	66	4.68	52	3.68
tense	67	27.55	59	31.89	33	15.22	15	6.46	11	4.61

Adatok a HuComTech korpusból

- Mondatszerkezet + beszéddallam
- F0-mozgás a tagmondat végén ($\leq 350\text{ms}$)

syntactic type	fall #	fall (%)	rise #	rise (%)	stagn. #	stagn. (%)	descend. #	descend. (%)	upward #	upward (%)
Type 1 complete	297	33	280	31.11	181	20.11	82	9.11	60	6.67
Type 13 incomplete	20	20	34	34	28	28	9	9	9	9
Type 14 sentence word	966	31.72	717	23.55	626	20.56	395	12.97	341	11.2

Adatok a HuComTech korpusból

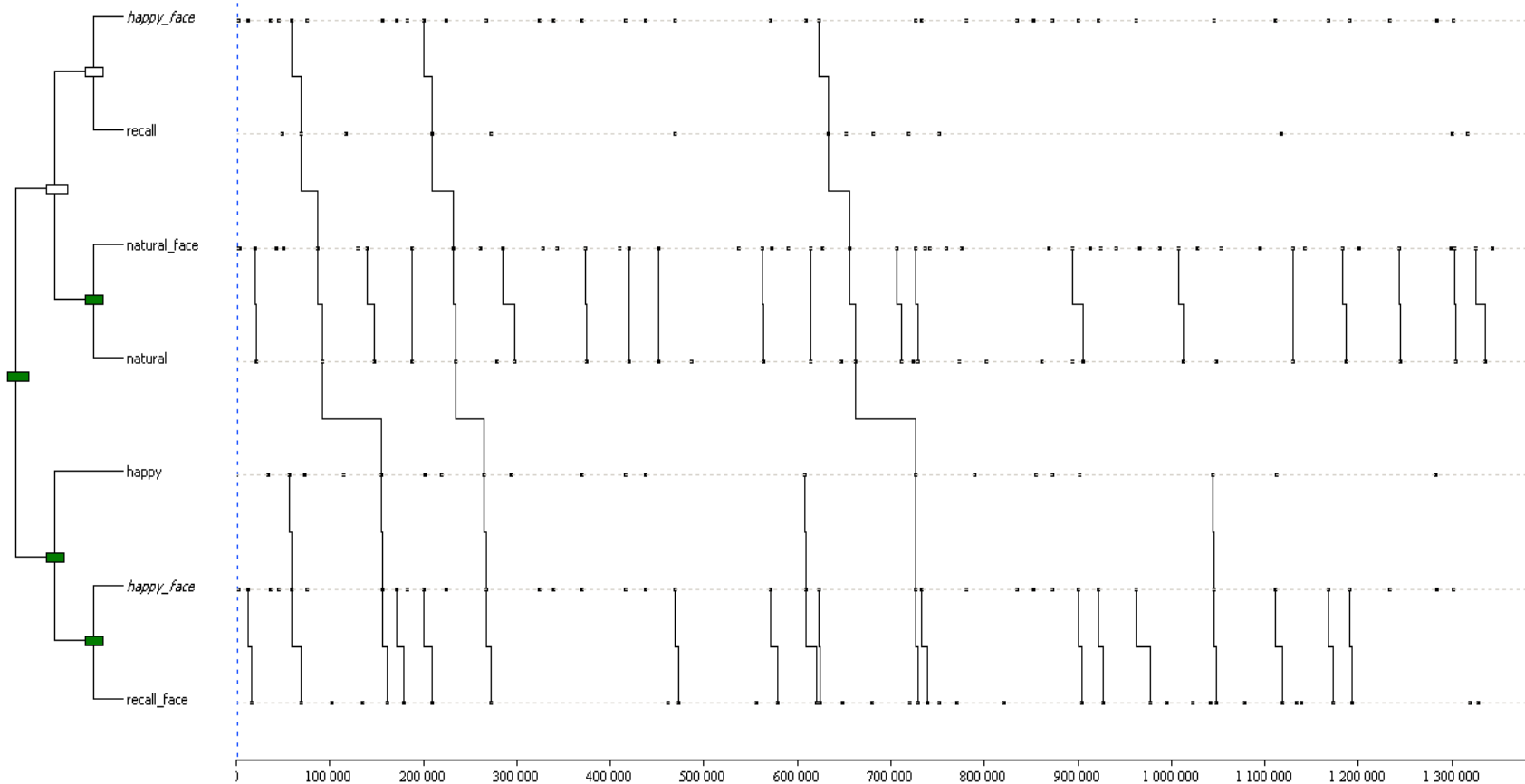
- Mondatszerkezet + beszéddallam
- Határozó: F0-mozgás a mondat elején

adverbial	fall #	fall (%)	rise #	rise (%)	stagn. #	stagn. (%)	descend. #	descend. (%)	upward #	upward (%)
hát	323	33.3	178	18.35	311	32.06	95	9.79	63	7
talán	3	25	6	50	3	25	0	8.33	0	8.33
de	160	35.63	144	32.07	97	21.6	29	6.46	19	4.23
még	17	32.69	14	26.92	16	30.77	4	7.69	1	1.92
már	8	34.78	6	34.78	8	34.78	1	4.35	0	0
ugye	3	37.5	2	25	2	25	0	0	1	12.5

Adatok a HuComTech korpuszból

(((happy_face recall) (natural_face natural)) (happy (happy_face recall_face)))

ID=163 N=3 Length=7 Duration=280003 %Duration=20 Log10(P(I))=-7,34 Log10(P(N))=-18,09



A kihívás

- A kommunikáció mögöttes időstruktúrájának a feltárása



Hozzáférés a HuComTech adatbázishoz

- Az annotáció teljes anyagát (audió, videó, annotációk) 2015. első felében hozzáférhetővé kívánjuk tenni
- Személyiségi jogok védelme miatt egyes adatok csak kutatás számára elérhetők
- MTA (Budapest)
- TLA (Nijmegen)



Köszönet munkatársainknak!

Abuczki Ágnes
Balázs László
Bertók Kornél
Bódog Alexa
Borbély Anikó
Campbell, Nick
Cseh Miklós Zsolt
Csipkés László
Csúry István
Fazekas Árpád
Földesi András
Ivaskó Livia
Keczán Marianna
Kertész András
Kiss Hermina
Németh T. Enikő
Oravec Csaba

Patkó Petra
Pápay Kinga
Péter Szabina
Prószéky Gábor
Sajó Levente
Seppänen, Tapio
Sandler, Wendy
Short, Harold
Staudt Alexandra
Szaszák György
Szeghalmy Szilvia
Sztanó László
Schwarcz Tibor
Tóth Csilla
Váradi Tamás
Vicsi Klára
Winkler István

*és a National Instruments
Hungary szakembereinek:*

Ábrahám László
Nagy Gábor
Nagy Norbert
Némethi Péter





Köszönjük a figyelmet!

