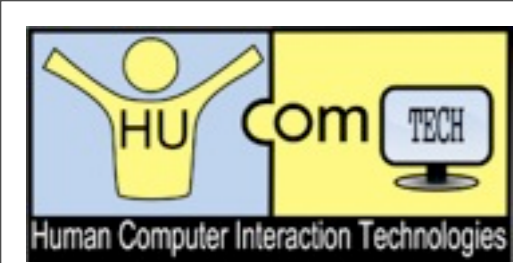




A HuComTech multimodális korpusz annotációja és adatbázisa

Hunyadi László
Debreceni Egyetem



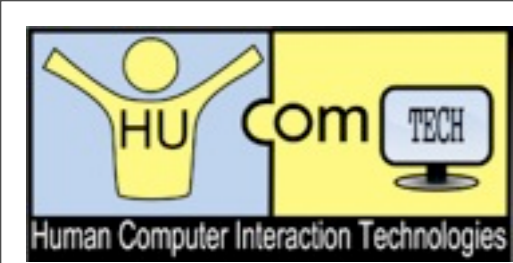
HuComTech:

Human-Computer Technologies

Hungarian Communication Technologies

az alábbi kutatócsoportok részvételével:

- számítógépes nyelvészet
- kommunikációelmélet
- pszichológia
- digitális képfeldolgozás
- mérnöki tudományok (robotika)



Néhány alapadat a korpuszról

A korpusz létrehozásának célja

az ember-ember kommunikáció azon elemeinek és szerkezeti viszonyainak a tanulmányozása, amelyek

- relevánsak az ember-gép kommunikációban (HCI-ben)
- technológiai szempontból megvalósíthatók

További cél, hogy

- megismerjük az emberi kommunikáció multimodális természetét (mind verbális, mind nem verbális vonatkozásait)
- leírjuk az emberi kommunikációt egy multimodális, holisztikus modellben



Szándékunk kellő mennyiségű adatot megfelelő formában nyújtani minimálisan a következő tudományterületek számára:

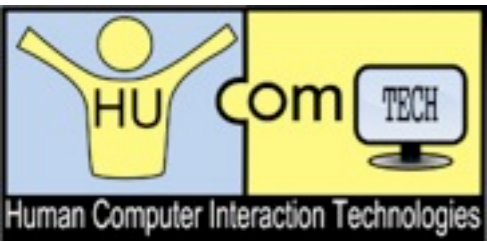
nyelvészet

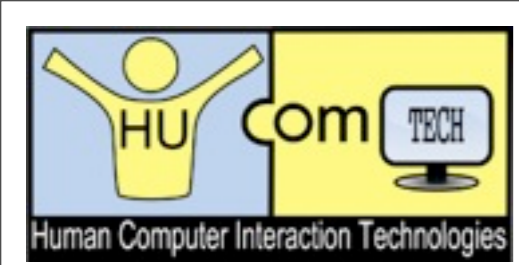
nyelvtechnológia (beszédfelismerő szoftverek támogatása)

viselkedépszichológia

robotika

és mások





A korpusz:

A korpusz:

- kb. 60 óra videofelvétel 111 beszélővel 18-29 éves kor között

A korpusz:

- kb. 60 óra videofelvétel 111 beszélővel 18-29 éves kor között
- \approx 450.000 szövegszó

A korpusz:

- kb. 60 óra videofelvétel 111 beszélővel 18-29 éves kor között
- ≈ 450.000 szövegszó felvételenként

A korpusz:

- kb. 60 óra videofelvétel 111 beszélővel 18-29 éves kor között
- \approx 450.000 szövegszó
felvételenként
- 15 felolvasott mondat

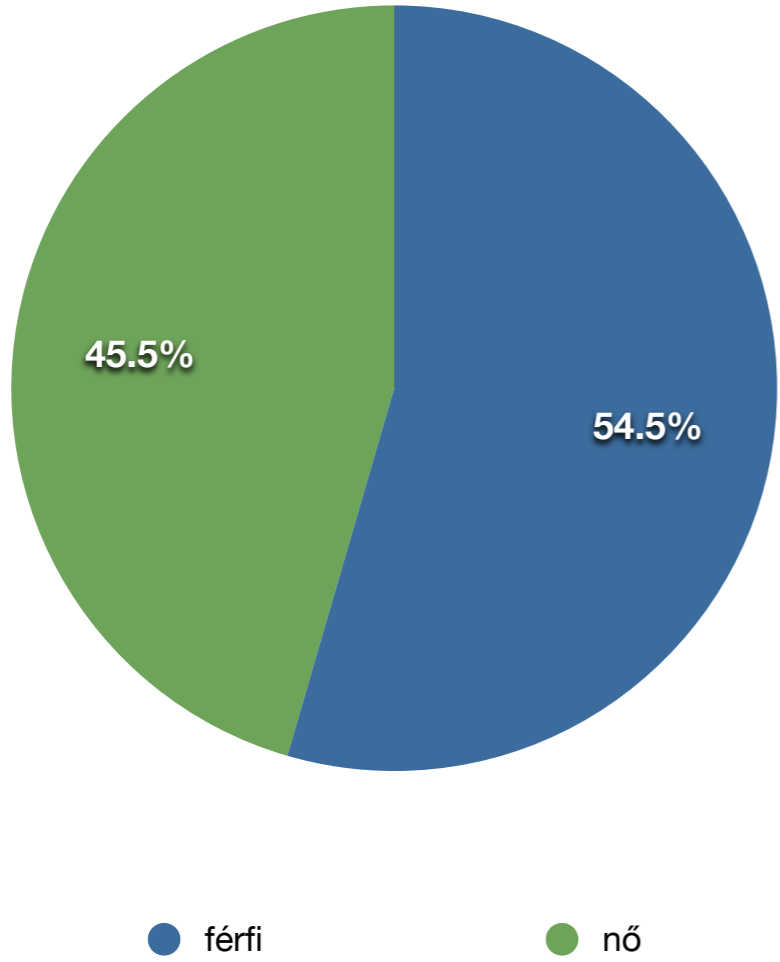
A korpusz:

- kb. 60 óra videofelvétel 111 beszélővel 18-29 éves kor között
- \approx 450.000 szövegszó
felvételenként
- 15 felolvasott mondat
- 10 perces irányított párbeszéd (állásinterjúk)

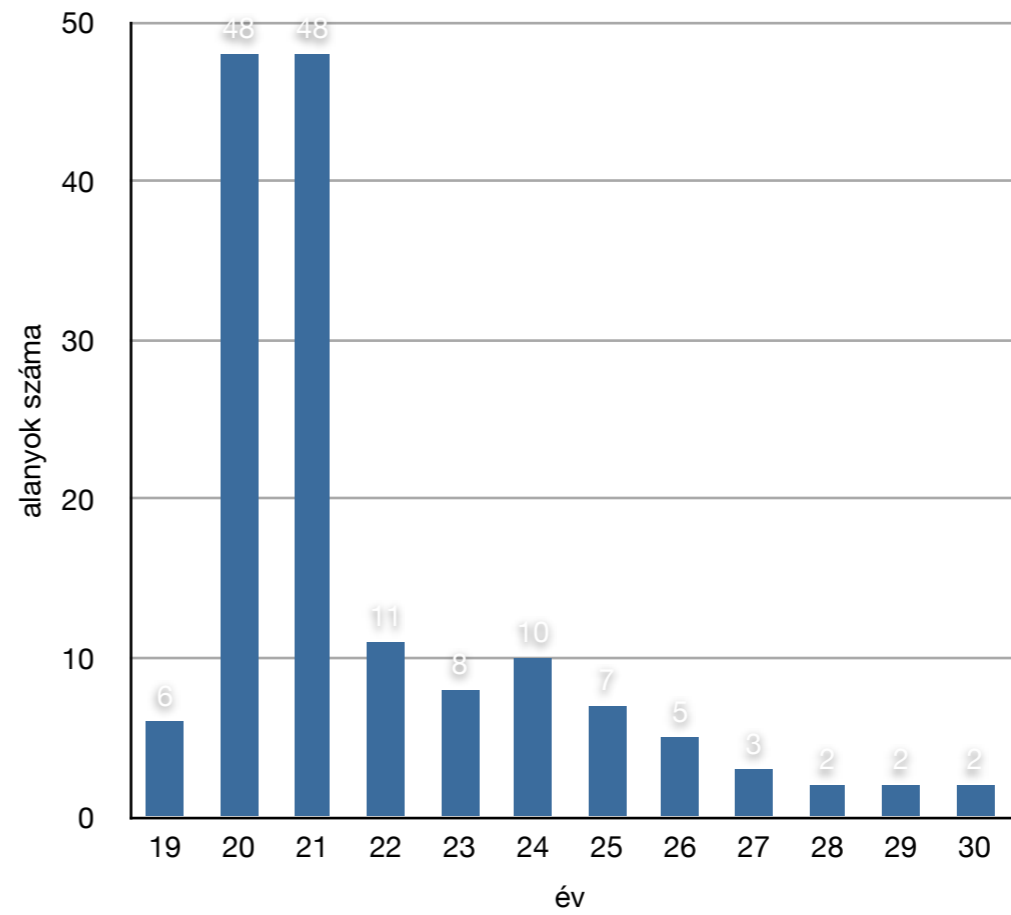
A korpusz:

- kb. 60 óra videofelvétel 111 beszélővel 18-29 éves kor között
- \approx 450.000 szövegszó felvételenként
- 15 felolvasott mondat
- 10 perces irányított párbeszéd (állásinterjúk)
- 15 perces kötetlen párbeszéd

Az alanyok nemek szerinti eloszlása



Az alanyok életkor szerinti eloszlása





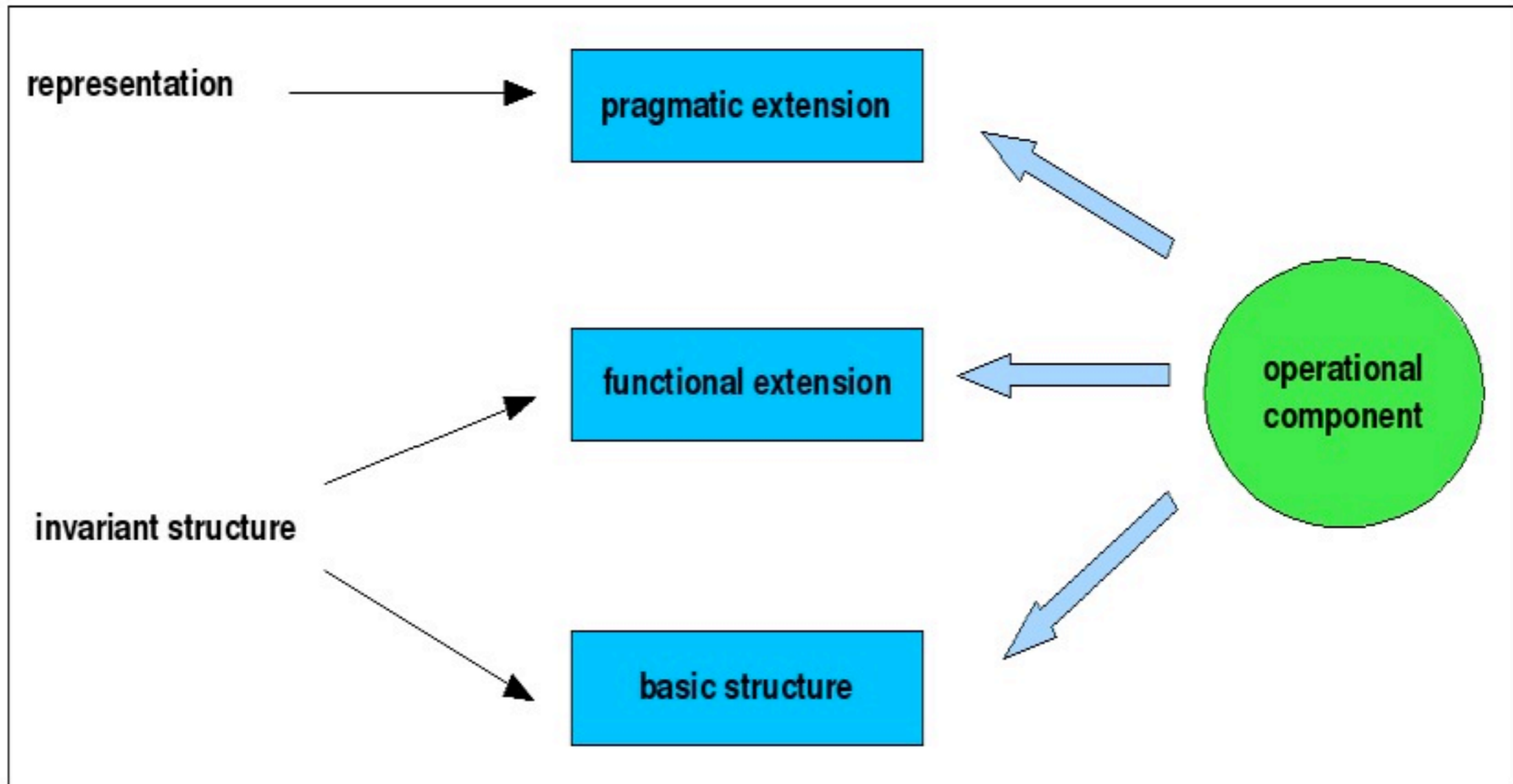
Az annotáció alapvető célja

az unimodális és multimodális markerek
annotálása és az időben egybeeső
markerek fúziója

Az annotálandó markereket a
kommunikáció egy elméleti-technológiai
modelljében határozzuk meg

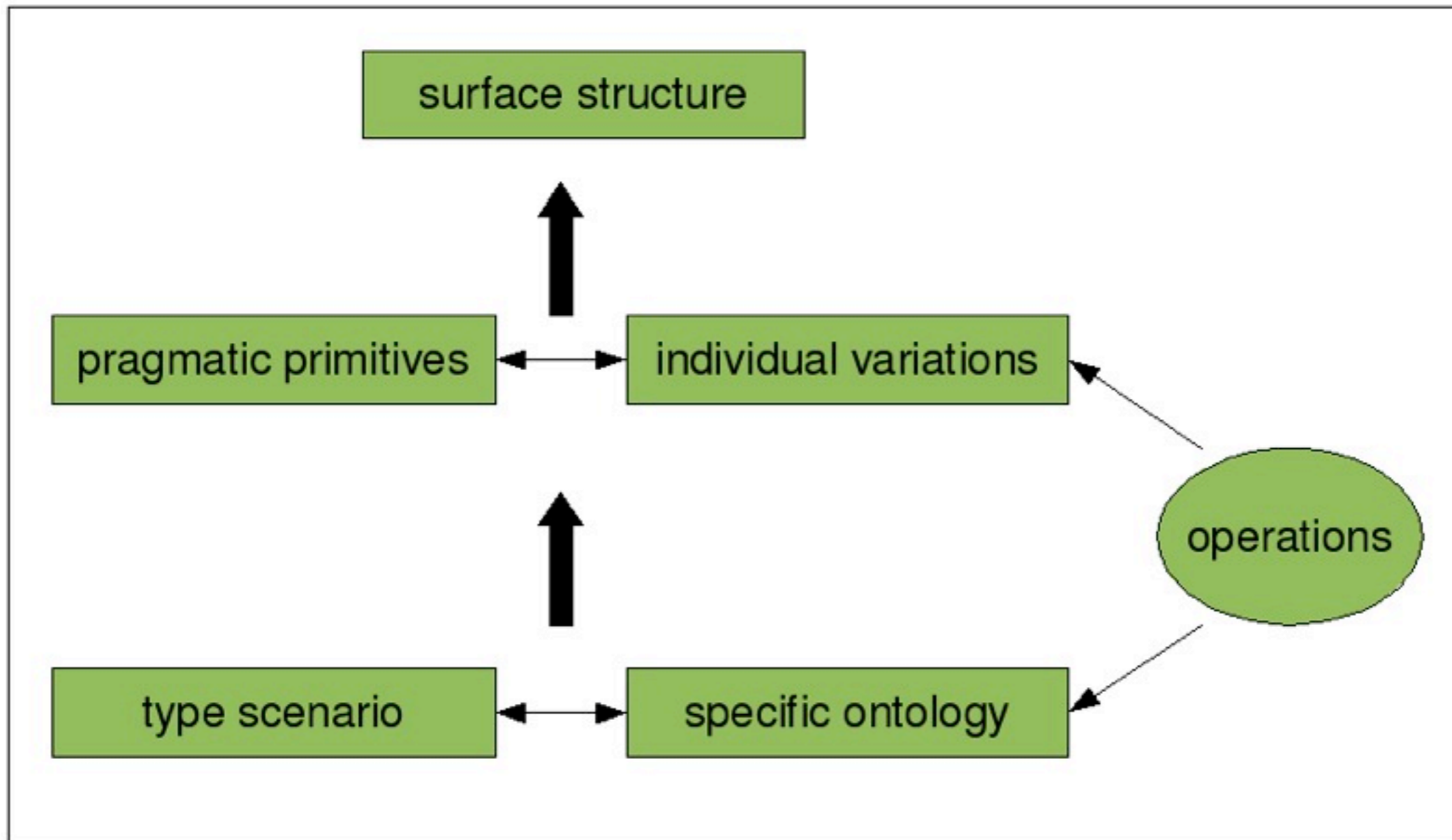
A modell alapszerkezete

Moduláris:



A modell alapszerkezete

A pragmatikai kiterjesztés



A kísérletek jellemzői

1. video

Annotáció

unimodális (videó, audió)
multimodális (videó + audió)

kézi
automatikus

a fizikai jellemzők leírása (különösen a videó esetében)
interpretatív annotálások

különös figyelemmel az érzelmek kifejezésére és a videó és az audió markerek multimodális egybeesésére
az annotáció fő területei:

pragmatika
szintaxis
prozódia

Az annotálás szintjei és címkéi

Audió

Az annotálás szintjei és címkéi

Audió

IP-szint: HC, SC, EM, IN, BC, HE, RE, IT, SL, V

Az annotálás szintjei és címkéi

Audió

IP-szint: HC, SC, EM, IN, BC, HE, RE, IT, SL, V

társalgás szintje: TT, TK, BC, SL

Az annotálás szintjei és címkéi

Audió

IP-szint: HC, SC, EM, IN, BC, HE, RE, IT, SL, V

társalgás szintje: TT, TK, BC, SL

érzelmek szintje: neutral, sad, happy/laughing, surprised, recall, tensed (és fokozataik: strong, moderate, reduced), other, silence

Az annotálás szintjei és címkéi

Videó

Az annotálás szintjei és címkéi

Videó

comevent: start, end

Az annotálás szintjei és címkéi

Videó

comevent: start, end

deictic: addressee, self, measure, object, shape; left, right, both

Az annotálás szintjei és címkéi

Videó

comevent: start, end

deictic: addressee, self, measure, object, shape; left, right, both

emblems: attention, agree, block, disagree, doubt, doubt-shrug, refusal, surprise, more-or-less, number, finger-ring, hands up, one hand other hand, other

Az annotálás szintjei és címkéi

Videó



Az annotálás szintjei és címkéi

Videó

emotions: natural, happy, recall, sad, surprise, tense; (and degrees of: strong, moderate, reduced)

Az annotálás szintjei és címkéi

Videó

emotions: natural, happy, recall, sad, surprise, tense; (and degrees of: strong, moderate, reduced)

headshift: lower, turn, raise, shake, nod; sideways, left, right

Az annotálás szintjei és címkéi

Videó

emotions: natural, happy, recall, sad, surprise, tense; (and degrees of: strong, moderate, reduced)

headshift: lower, turn, raise, shake, nod; sideways, left, right

touchmotion: hair, leg, arm, face, eye, ear, chin, mouth, neck, bust, forehead, nose, glasses; tap, scratch; left, right

Az annotálás szintjei és címkéi

Videó

Az annotálás szintjei és címkéi

Videó

posture: crossing arm, holding head, lean back, lean forward, lean left, lean right, rotate right, rotate left, shoulder up, upright

Az annotálás szintjei és címkéi

Videó

posture: crossing arm, holding head, lean back, lean forward, lean left, lean right, rotate right, rotate left, shoulder up, upright

handshape: breaking, fist, crossing fingers, open flat, open spread, thumb out, index out; left, right, both

Az annotálás szintjei és címkéi

Videó

Az annotálás szintjei és címkéi

Videó

facial expressions: natural, happy, recall, sad, surprise, tense (and degrees of: moderate, reduced, strong)

Az annotálás szintjei és címkéi

Videó

facial expressions: natural, happy, recall, sad, surprise, tense (and degrees of: moderate, reduced, strong)

eyebrows: scowl, up; left, right, both

Az annotálás szintjei és címkéi

Videó

facial expressions: natural, happy, recall, sad, surprise, tense (and degrees of: moderate, reduced, strong)

eyebrows: scowl, up; left, right, both

gaze: blink, up, down, left, right, forwards, left-up, left-down, right-up, right-down

Az annotálás szintjei és címkéi

Szintaxis

strukturális szegmentálás:

Az annotálás szintjei és címkéi

Szintaxis

strukturális szegmentálás:

tagmondathatárok

Az annotálás szintjei és címkéi

Szintaxis

strukturális szegmentálás:

tagmondathatárok

tagmondatok hierarchiája

Az annotálás szintjei és címkéi

Szintaxis

strukturális szegmentálás:

tagmondathatárok

tagmondatok hierarchiája

tagmondatok belső szerkezete (különös tekintettel a hiányzó elemekere)

Az annotálás szintjei és címkéi

Szintaxis és prozódia

(tagmondatok prozódiaja)

Az annotálás szintjei és címkéi

Szintaxis és prozódia

(tagmondatok prozódiaja)

pitch movement: rise, fall, stagnant + finomabb osztályozás

Az annotálás szintjei és címkéi

Szintaxis és prozódia

(tagmondatok prozodiája)

pitch movement: rise, fall, stagnant + finomabb osztályozás

intensity: increase, decrease, stagnant + finomabb osztályozás

Az annotálás szintjei és címkéi

Szintaxis és prozódia

(tagmondatok prozodiája)

pitch movement: rise, fall, stagnant + finomabb osztályozás

intensity: increase, decrease, stagnant + finomabb osztályozás

pause/duration: increase, decrease, stagnant + finomabb osztályozás



Az annotálás szintjei és címkéi

Pragmatika - multimodális



Az annotálás szintjei és címkéi

Pragmatika - multimodális

Annotáció: DiAMSL a szövegalapú eseményekre, az audió számos szintjén (pl. fordulóváltás, társalgás)



Az annotálás szintjei és címkéi

Pragmatika - multimodális

Annotáció: DiAMSL a szövegalapú eseményekre, az audió számos szintjén (pl. fordulóváltás, társalgás)

Multimodalitás - az audió + videó komplex egységében multimodális kommunikatív esemény



Az annotálás szintjei és címkéi

Pragmatika - multimodális



Az annotálás szintjei és címkéi

Pragmatika - multimodális

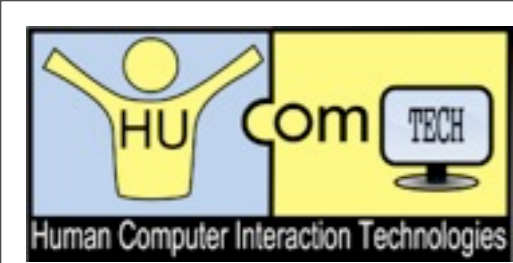
kommunikatív esemény típusok: constatives directives, comissives, acknowledgements, egyik sem

Az annotálás szintjei és címkéi

Pragmatika - multimodális

kommunikatív esemény típusok: constatives directives, comissives, acknowledgements, egyik sem

kommunikatív események támogatása: backchannel, politeness markers, corrections, nincs támogatás



Az annotálás szintjei és címkéi

Pragmatika - multimodális

Az annotálás szintjei és címkéi

Pragmatika - multimodális

tematikus kontroll: topic initiation, elaboration, topic change (contextual, non-contextual)



Az annotálás szintjei és címkéi

Pragmatika - multimodális

tematikus kontroll: topic initiation, elaboration, topic change (contextual, non-contextual)

információszerkezet: adott vs. új információ



Az annotálás szintjei és címkéi

Pragmatika - unimodális



Az annotálás szintjei és címkéi

Pragmatika - unimodális

agreement: uninterested, disagree, block, uncertainty; full, partial



Az annotálás szintjei és címkéi

Pragmatika - unimodális

agreement: uninterested, disagree, block, uncertainty; full, partial

attention: calling, paying

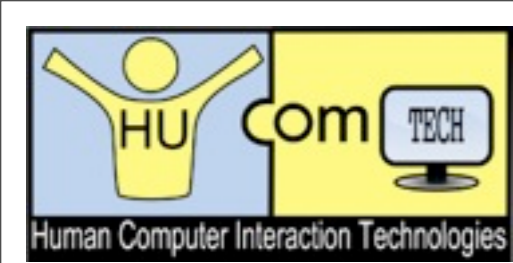
Az annotálás szintjei és címkéi

Pragmatika - unimodális

agreement: uninterested, disagree, block, uncertainty; full, partial

attention: calling, paying

deixis



Az annotálás szintjei és címkéi

Pragmatika - unimodális



Az annotálás szintjei és címkéi

Pragmatika - unimodális

information: received novelty



Az annotálás szintjei és címkéi

Pragmatika - unimodális

information: received novelty

turn-management: intending to start speaking, start speaking successfully, end speaking, breaking in

Példák a multimodális együttjárásra

Fordulóváltás

Példák a multimodális együttjárásra

Fordulóváltás

turn-give: forward, blink, down, left-down, right-down

Példák a multimodális együttjárásra

Fordulóváltás

turn-give: forward, blink, down, left-down, right-down

turn-take: forward, blink, down, left-down, right-down

Példák a multimodális együttjárásra

Fordulóváltás

turn-give: forward, blink, down, left-down, right-down

turn-take: forward, blink, down, left-down, right-down

break-in_turn-keep: forwards, blink, up, down, left-down, right-down

Példák a multimodális együttjárásra

Érzelmek vs. gesztusok

Példák a multimodális együttjárásra

Érzelmek vs. gesztusok

A bizonytalanság 'uncertainty' leginkább a nyitott, kitért 'open spread' kézi gesztussal jár együtt, ritkábban az összefonott kézzel 'crossing fingers'.

Példák a multimodális együttjárásra

Érzelmek vs. gesztusok

A bizonytalanság 'uncertainty' leginkább a nyitott, kitért 'open spread' kézi gesztussal jár együtt, ritkábban az összefonott kézzel 'crossing fingers'.

Az egyetértés 'agreement' ugyancsak e két kézi gesztussal jár együtt.

Példák a multimodális együttjárásra

Érzelmek vs. gesztusok

A bizonytalanság 'uncertainty' leginkább a nyitott, kitarított 'open spread' kézi gesztussal jár együtt, ritkábban az összefonott kézzel 'crossing fingers'.

Az egyetértés 'agreement' ugyancsak e két kézi gesztussal jár együtt.

A kétség 'doubt' a fentiekén kívül együtt jár még az oldalrafordulással 'sideways' is.

Példák a multimodális együttjárásra

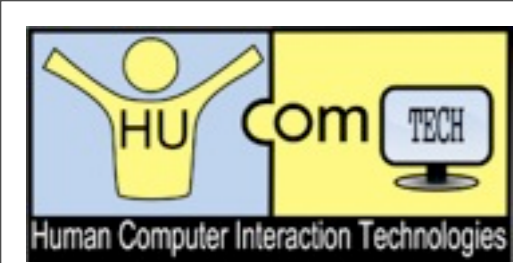
Érzelmek vs. gesztusok

Ez tehát azt mutatja, hogy a gesztusok egyértelműsítését is multimodálisan lehet megtenni.

Videó annotáció (kézi vs. automatikus)

annotáció módszere	előny/hátrány (+/-)	
	fizikai értékek	interpretatív értékek
kézi	-	+
automatikus	+	-

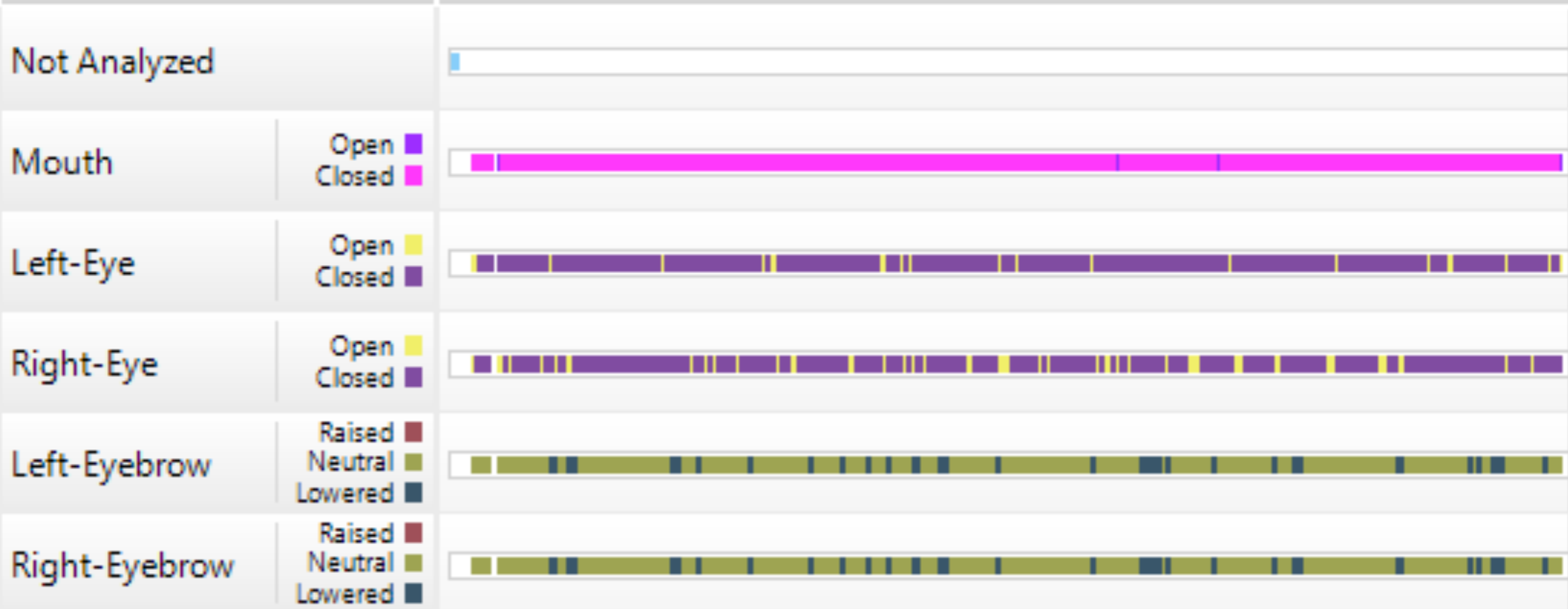
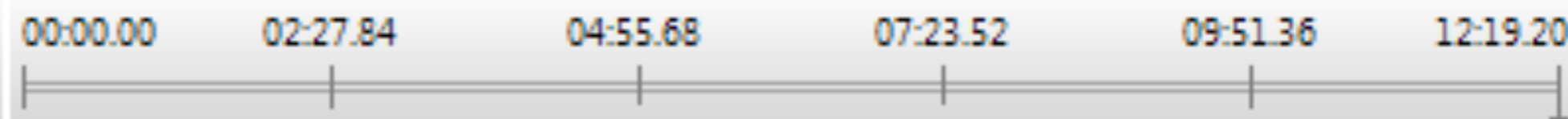
Lényegi különbségük: az automatikus annotáció ‘digitális’ képkockánkénti értékelés egy előre meghatározott méretű képsorban, míg a kézi annotálás ‘analóg’



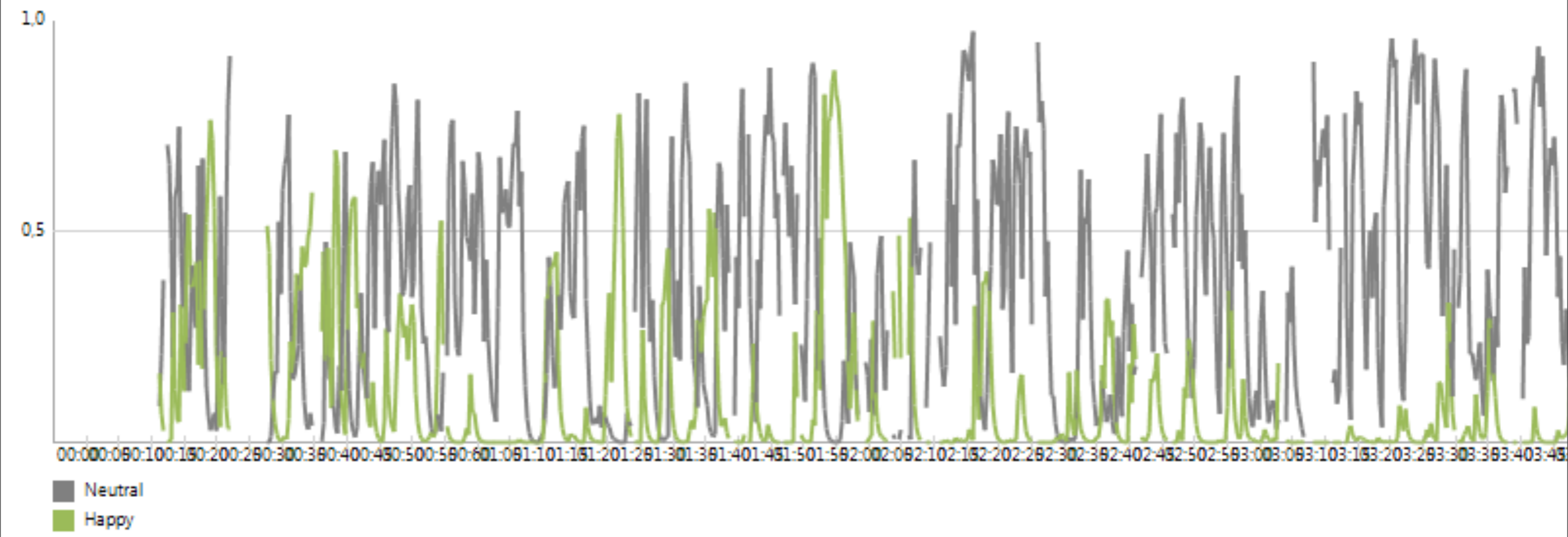
Példa az automatikus képi
videoannotálásra:

Noldus FaceReader

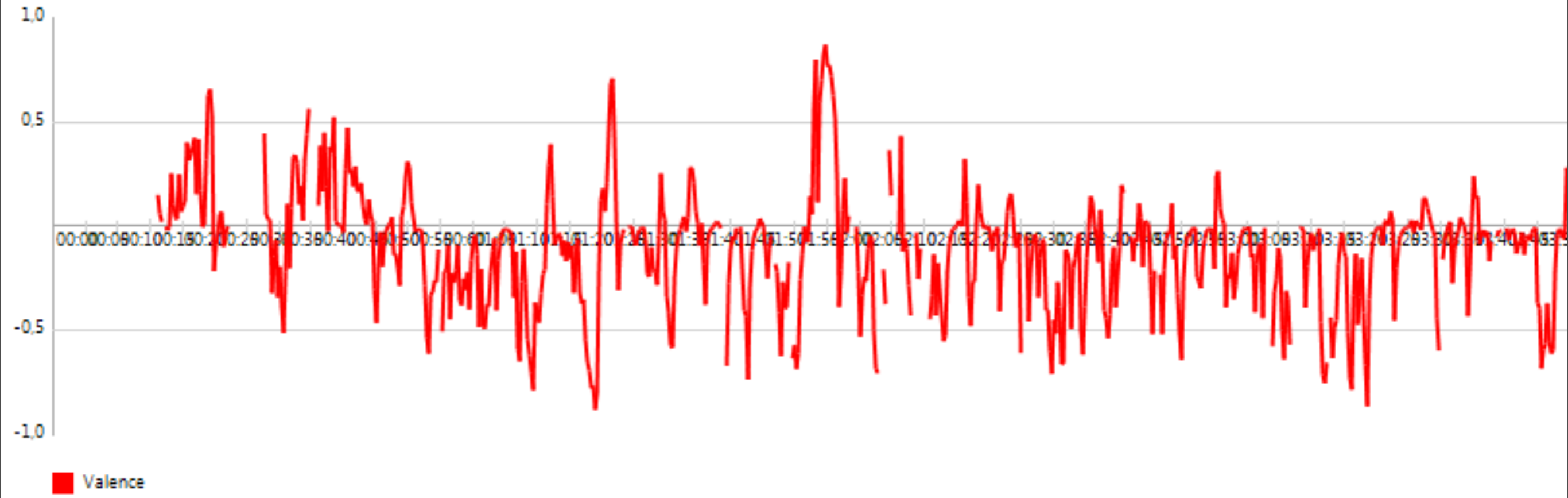
Bináris értékadás



Idővonal: natural vs. happy

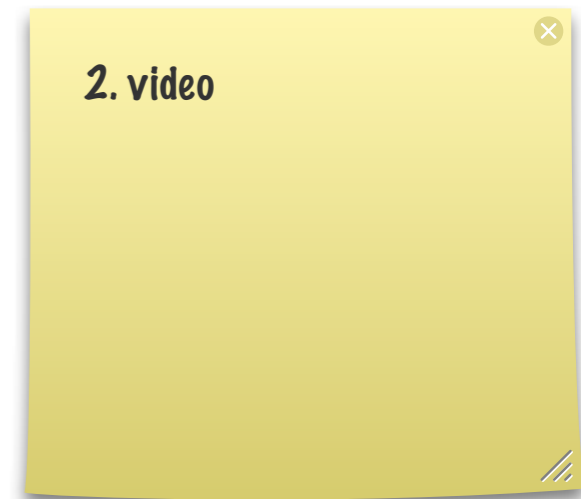
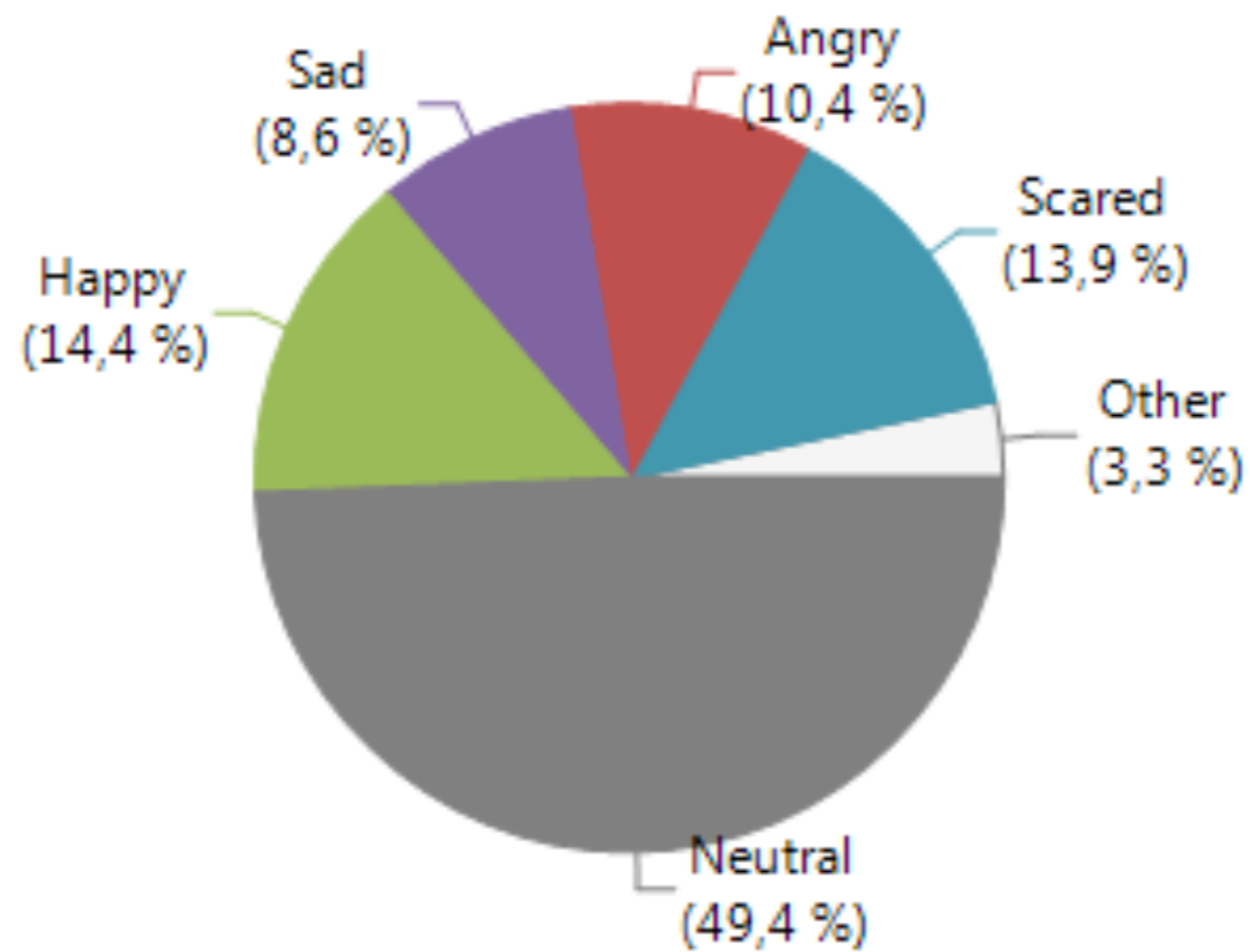


Valencia



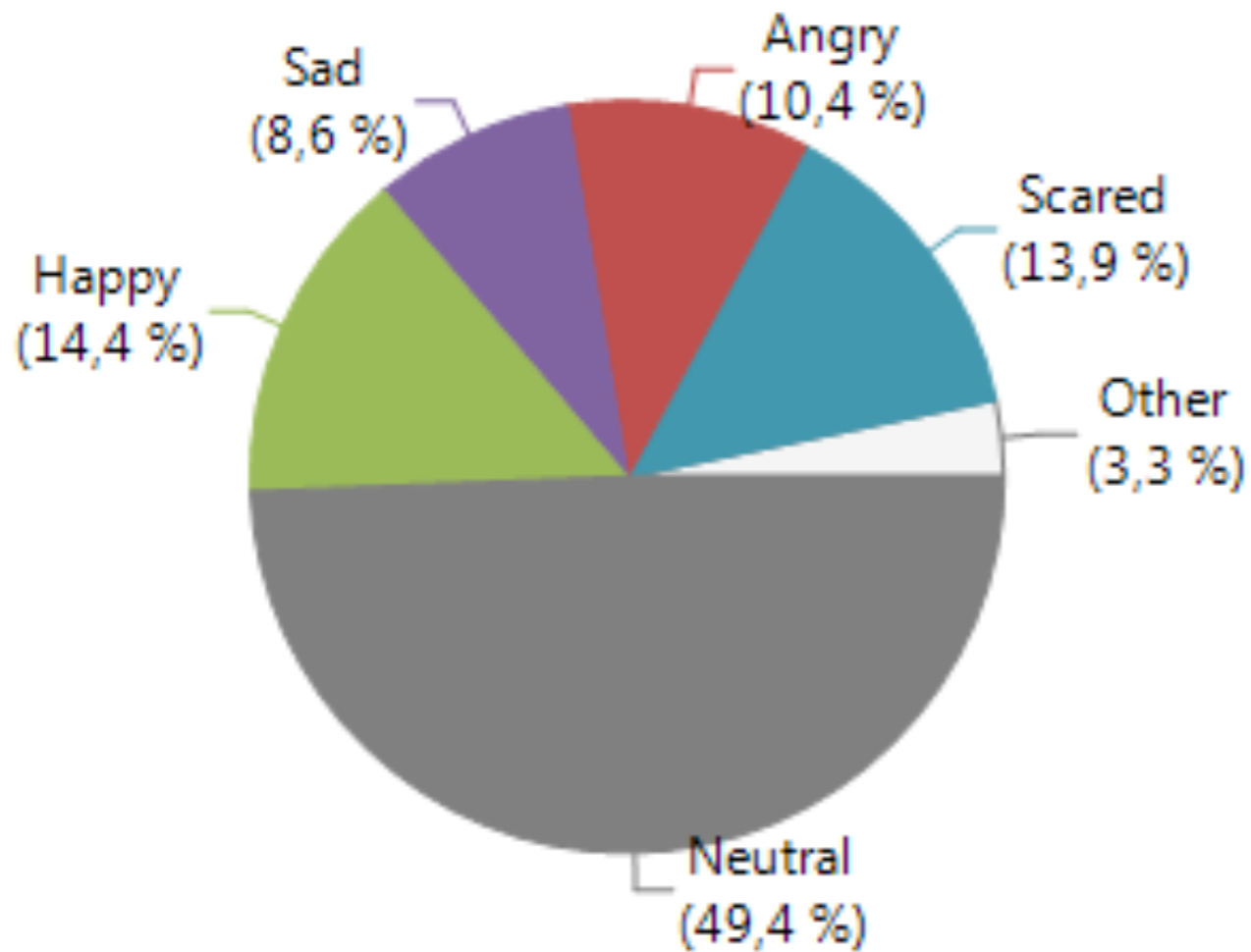
Összefoglaló statisztika

Analysis Block 1

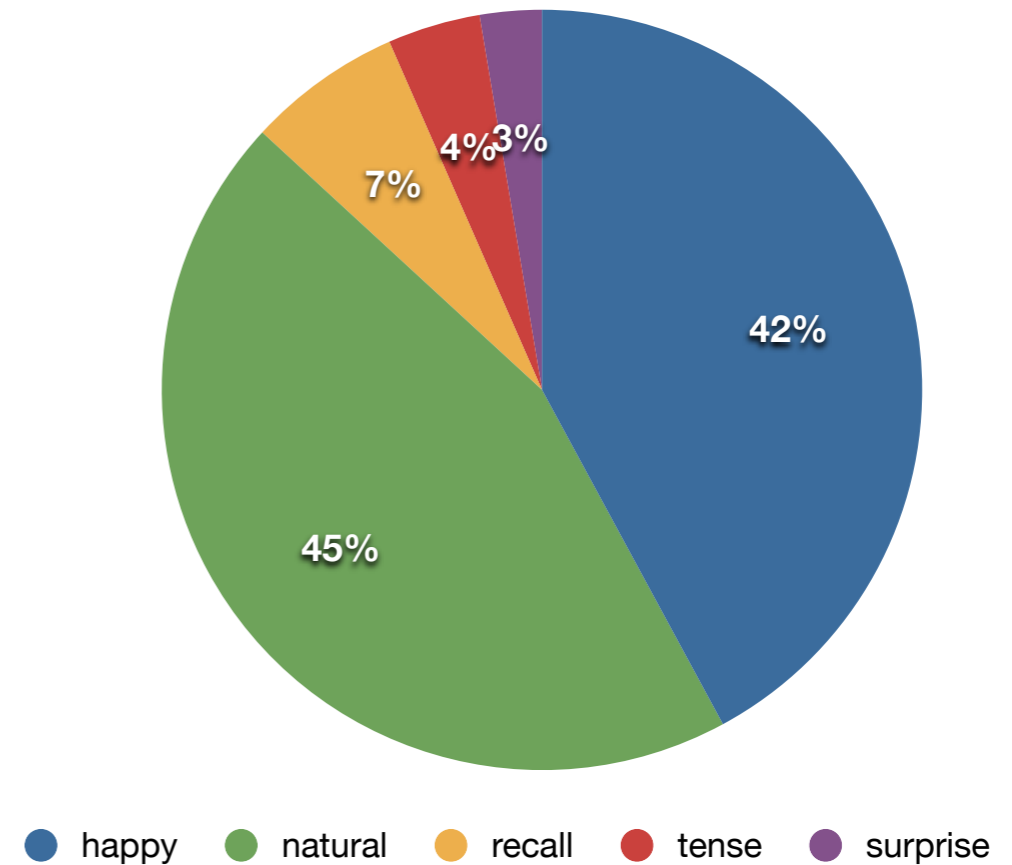


Összevetés: automatikus vs. kézi érzelemfelismerés

Analysis Block 1



Manual annotation



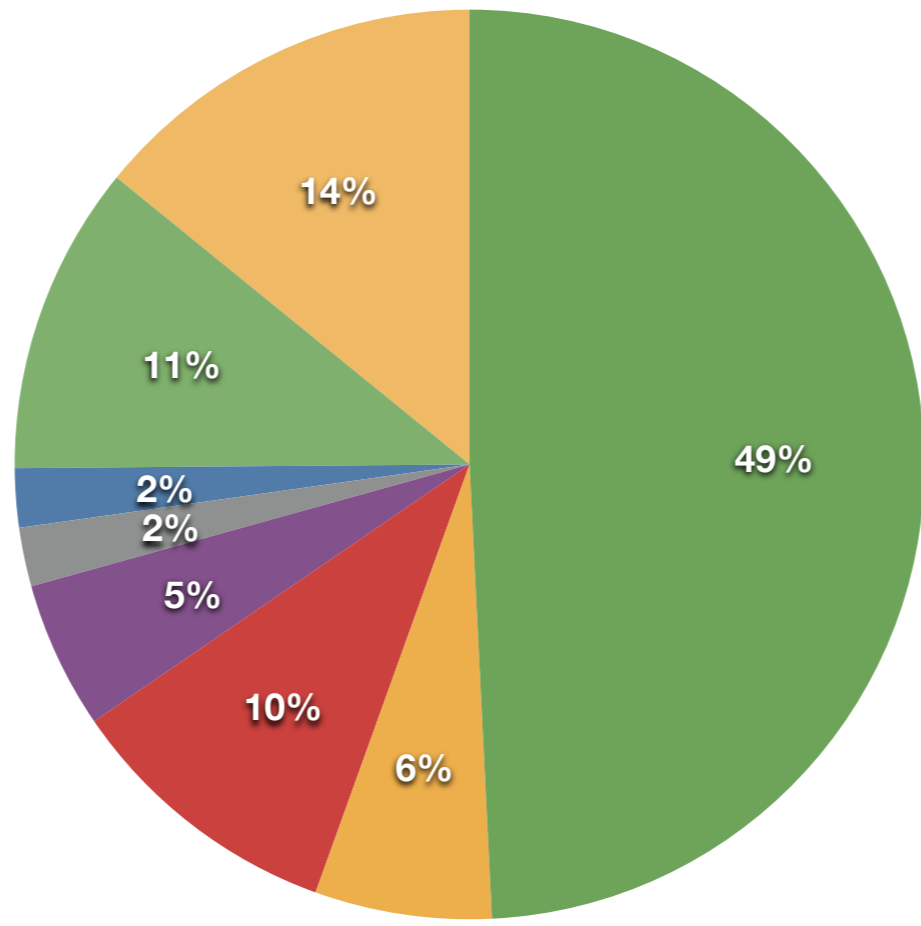
A beszélt nyelv szintaxisának annotálása és viszonya a prozódiahoz a *HuComTech* korpuszban

- Cél: nyelvtechnológia (beszéd > szöveg)
- kommunikációkutatás (a multimodális markerek együttjárása a kommunikációs események és az érzelmek jelölésére)
- nyelvtudomány (a szintaxis-prozódia interfész kutatása)

Type of missing element	Type of missing element according to syntactic code			
	Informal dialogs	%	Formal dialogs	%
1. nothing missing	2664	35.59	758	34.6
2. main clause	37	0.49	15	0.69
3. preceding clause	58	0.77	6	0.27
4. relative pronoun	89	1.19	22	1.01
5. conjunction	22	0.29	4	0.18
6. subject (grammatical)	3178	42.37	1167	54.14
7. subject (logical)	274	3.66	113	5.17
8. predicate	214	2.87	72	3.29
9. object	102	1.36	45	2.06
10. adverb	11	0.15	4	0.18
11. attribute	0	0	0	0
12. verb	10	0.13	0	0
13. unfinished clause	728	9.7	167	13.1
14 missing element not relevant	3375	45.05	769	35.21
			142.62	149.9

Type of missing element by frequency				
Type of missing element	Informal dialogs	%	Formal dialogs	%
14 missing element not relevant	3375	45.05	769	35.21
6. subject (grammatical)	3178	42.37	1167	54.14
1. nothing missing	2664	35.59	758	34.6
13. unfinished clause	728	9.7	167	13.1
7. subject (logical)	274	3.66	113	5.17
8. predicate	214	2.87	72	3.29
9. object	102	1.36	45	2.06
4. relative pronoun	89	1.19	22	1.01
3. preceding clause	58	0.77	6	0.27
2. main clause	37	0.49	15	0.69
5. conjunction	22	0.29	4	0.18
10. adverb	11	0.15	4	0.18
12. verb	10	0.13	0	0
11. attribute	0	0	0	0
Sum:		142.62		149.9

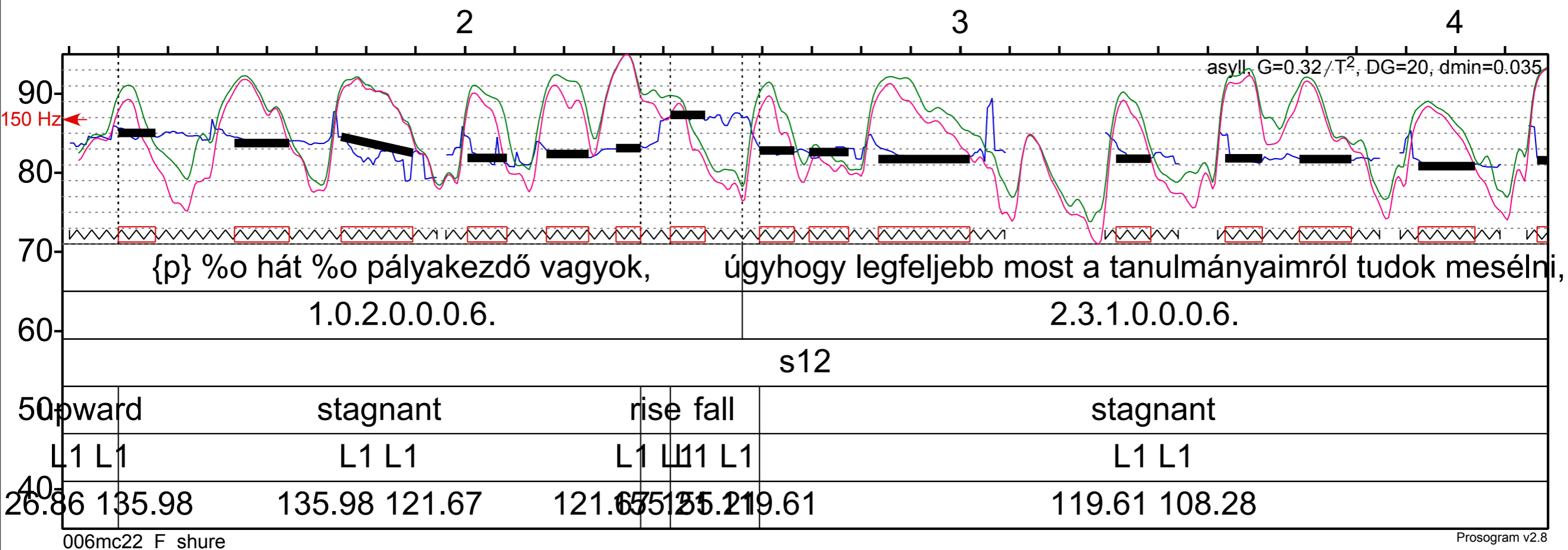
Befejezetlen tagmondat + szünethiány + tekintet



- informal: (blue)
- right-up (purple)
- blink (yellow)
- forwards (light green)
- left-up (grey)
- up (blue)
- left-down (orange)
- right-down (red)
- down (green)

Prozódiaannotálás a HuComTech korpuszban

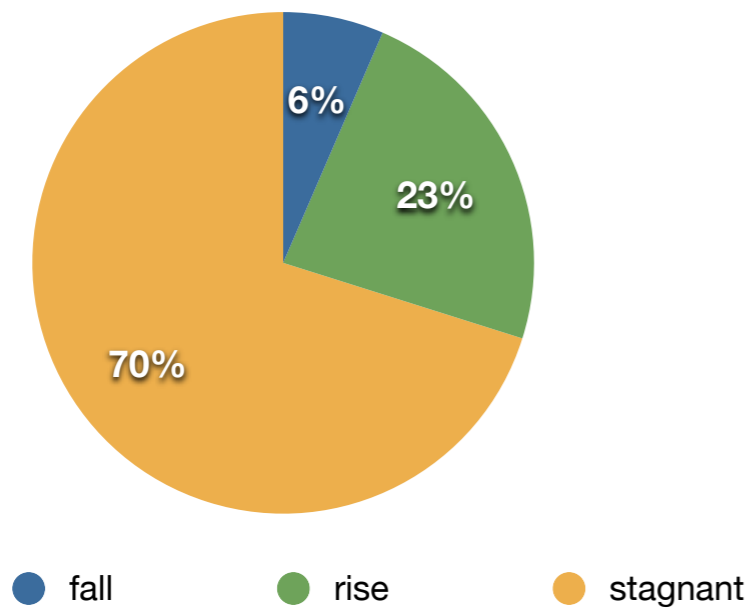
- “IP-szint” (F0-kontúr és szünet alapján, manuálisan)
- F0-kontúr (automatikus)
- intenzitásváltozás (automatikus, fejlesztés alatt)
- hangsúlydetektálás (automatikus, fejlesztés alatt)



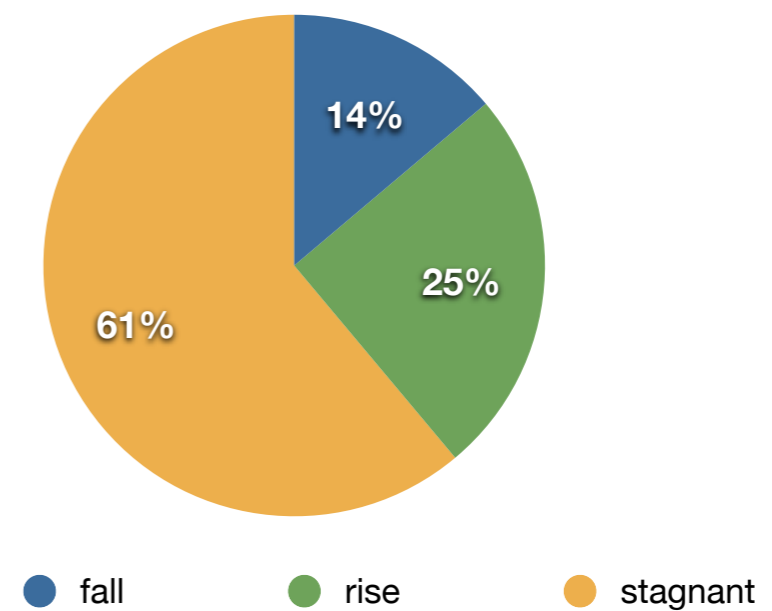
A számítások eredményeit összekapcsoljuk a szintaktikai típussal, de maga a számítás attól független

Szintaktikai típus vs. F0

Befejezetlen tagmondat (13-as típus), formális beszélgetés

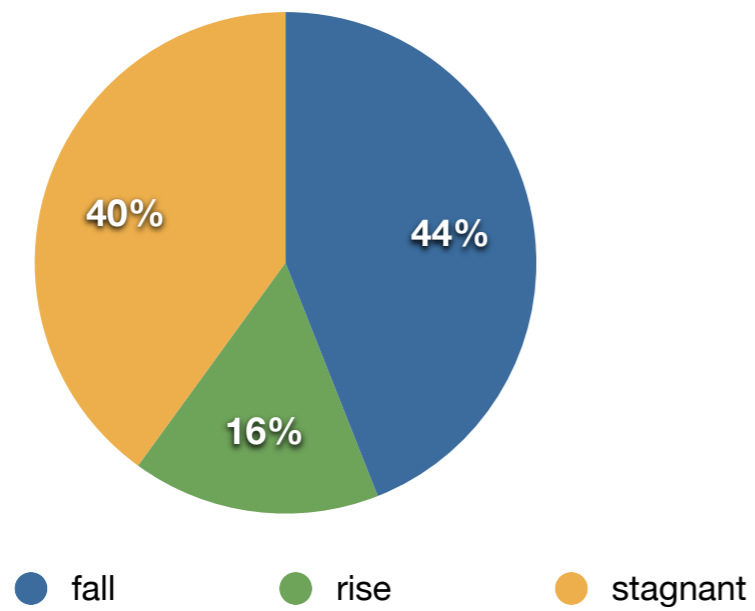


Befejezetlen tagmondat (13-as típus), informális beszélgetés

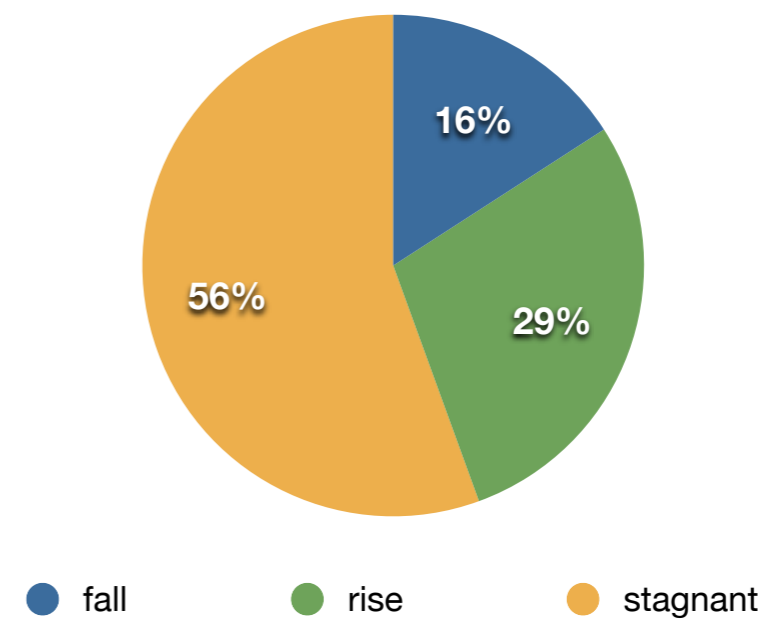


Szintaktikai típus vs. F0

Főmondat hiányzik (2-es típus), formális beszélgetés

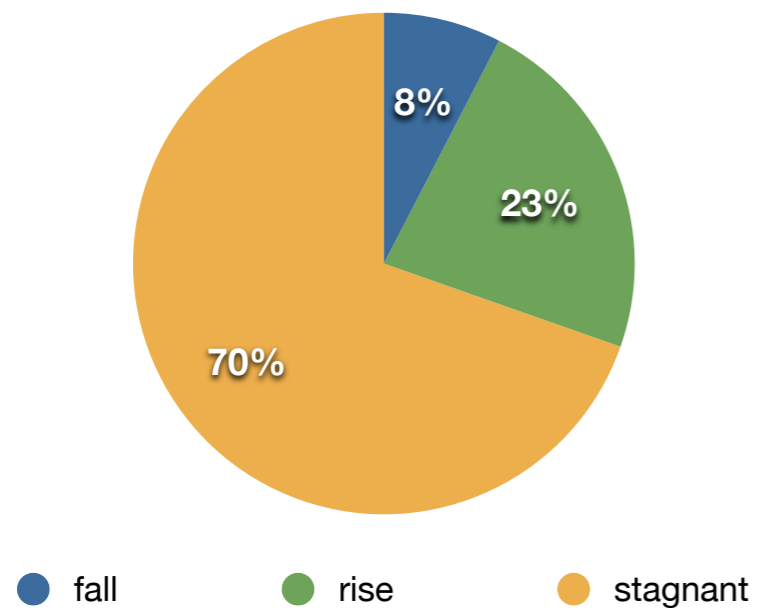


Főmondat hiányzik (2-es típus), iformális beszélgetés

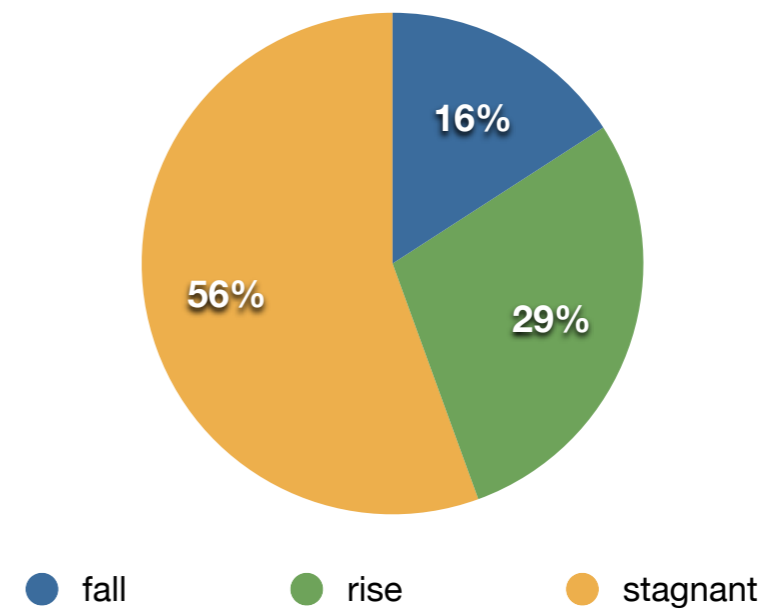


Szintaktikai típus vs. F0

Alárendelt tagmondat hiányzik (3-as típus), formális beszélgetés

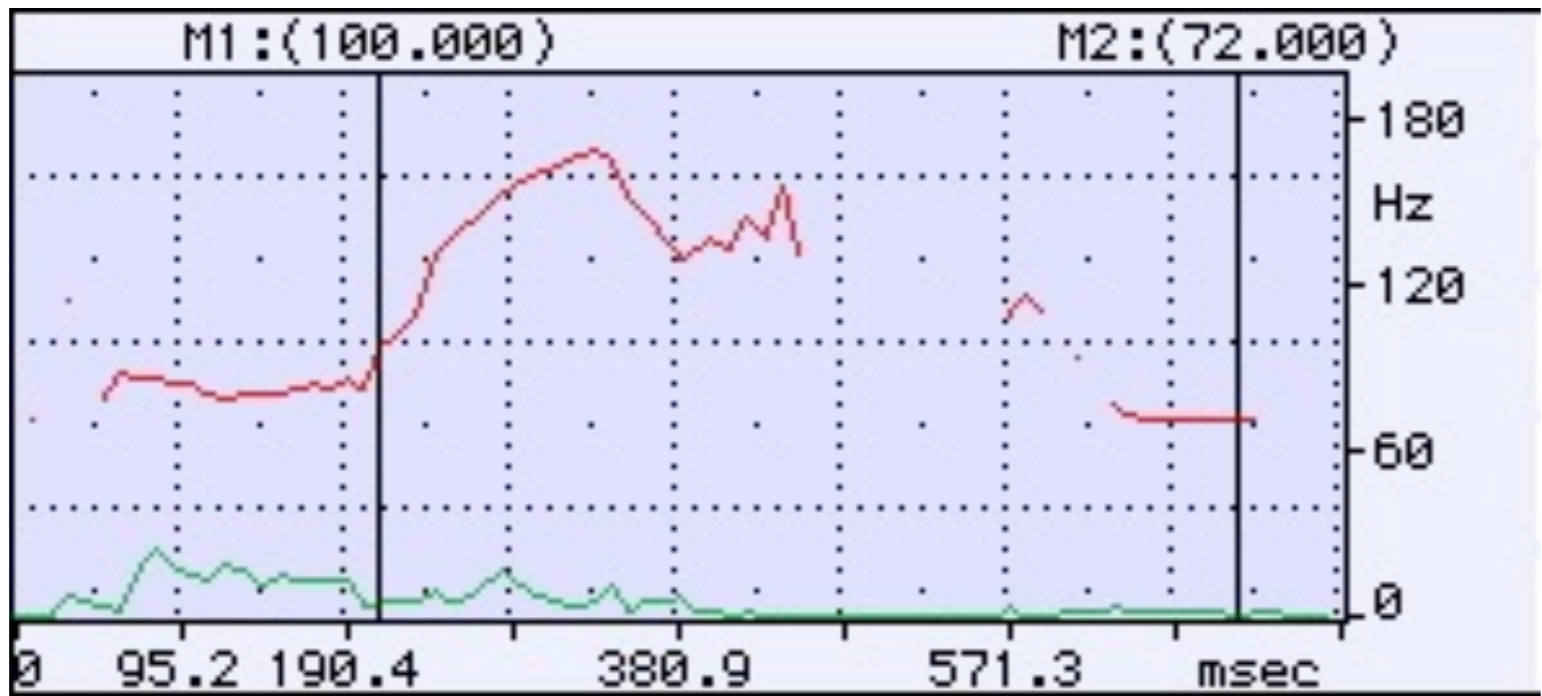


Alárendelt tagmondat hiányzik (3-as típus), formális beszélgetés



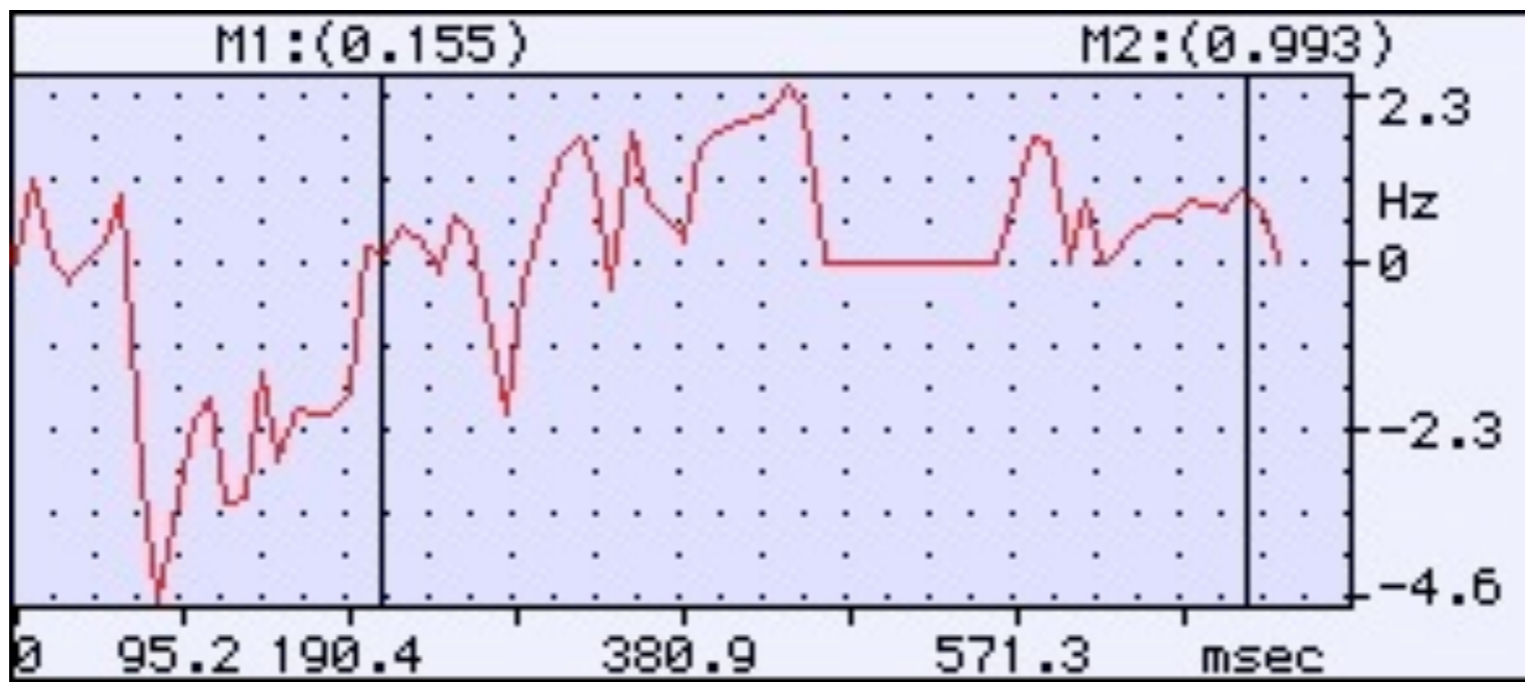
Hangsúlydetektálás

- alapja Hunyadi 2002
- PET: pitch and energy over time
- a hangsúly az F0 és az intenzitás interakciójának az eredménye: relatív prominencia
- abszolút PET-érték + időtartam

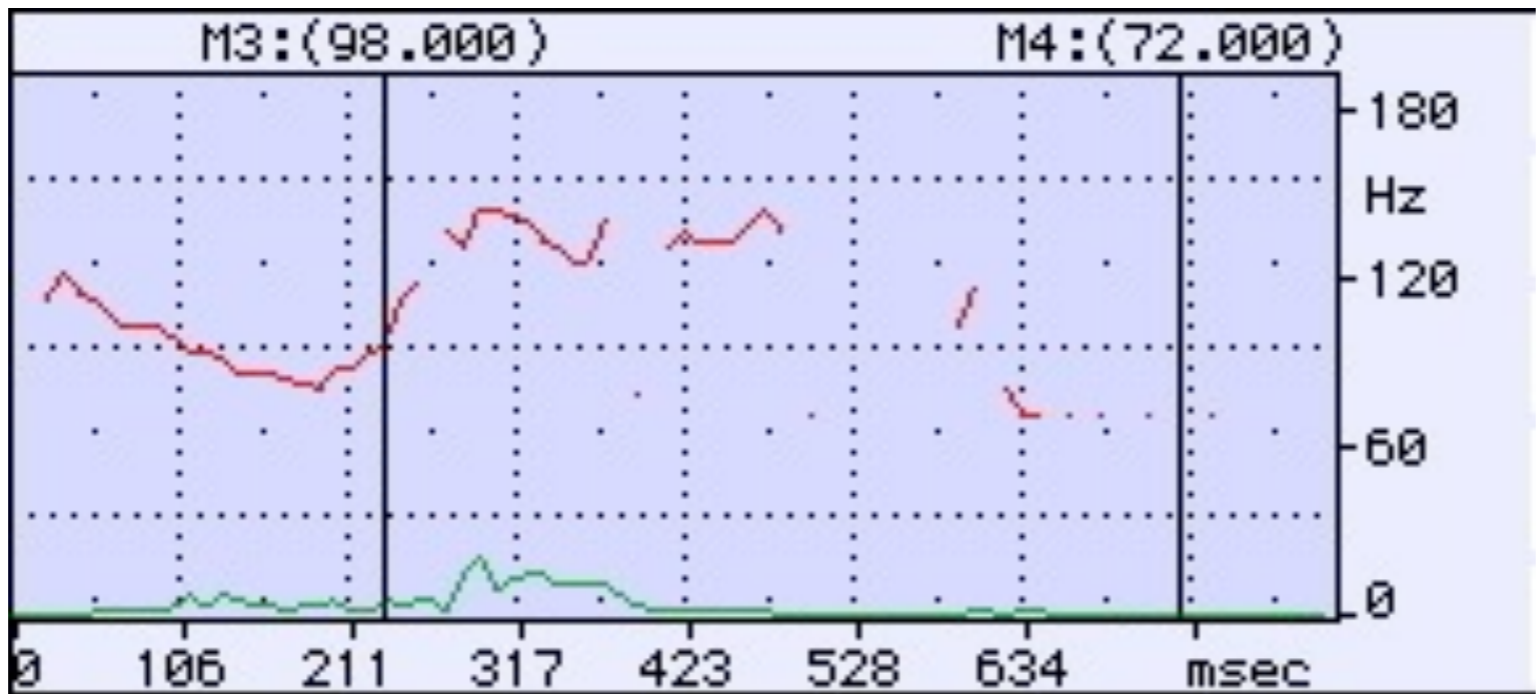


F0 vs. intenzitás

Ő látta?

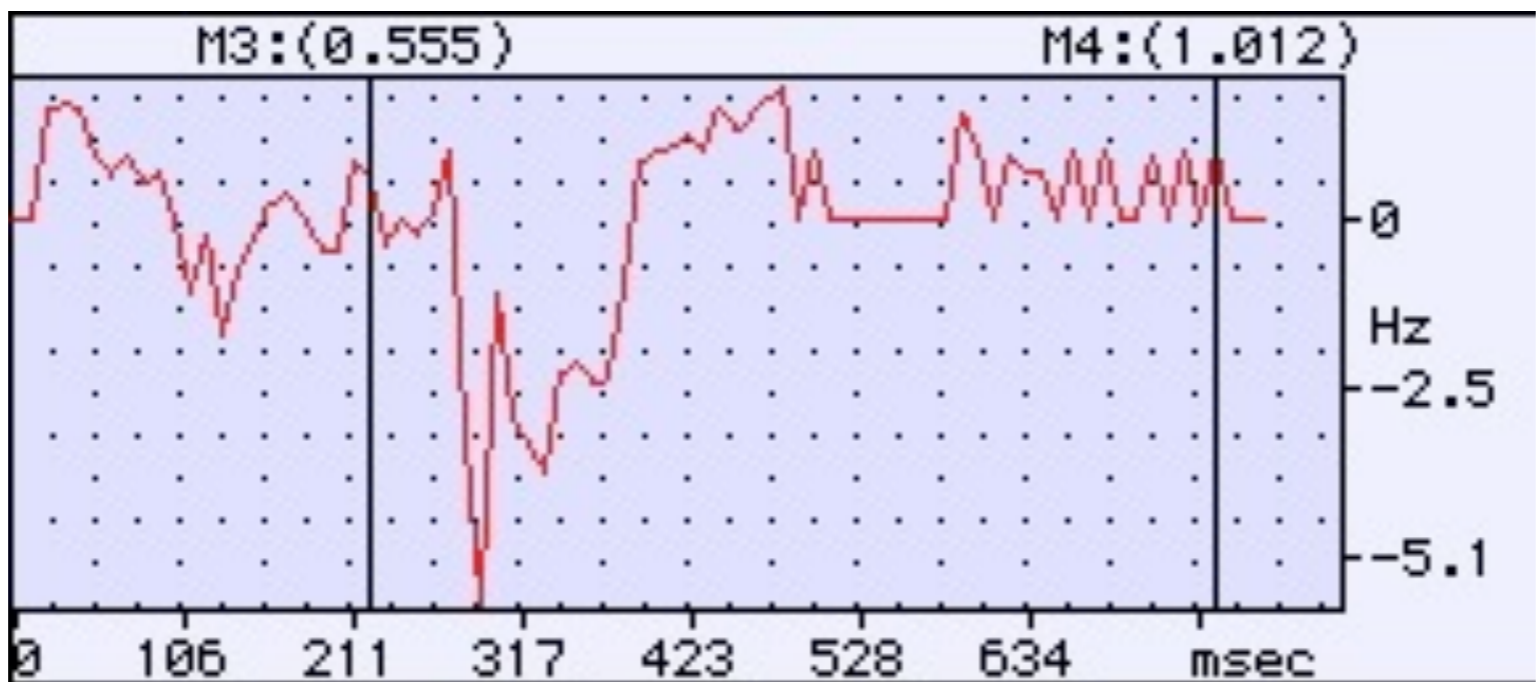


PET

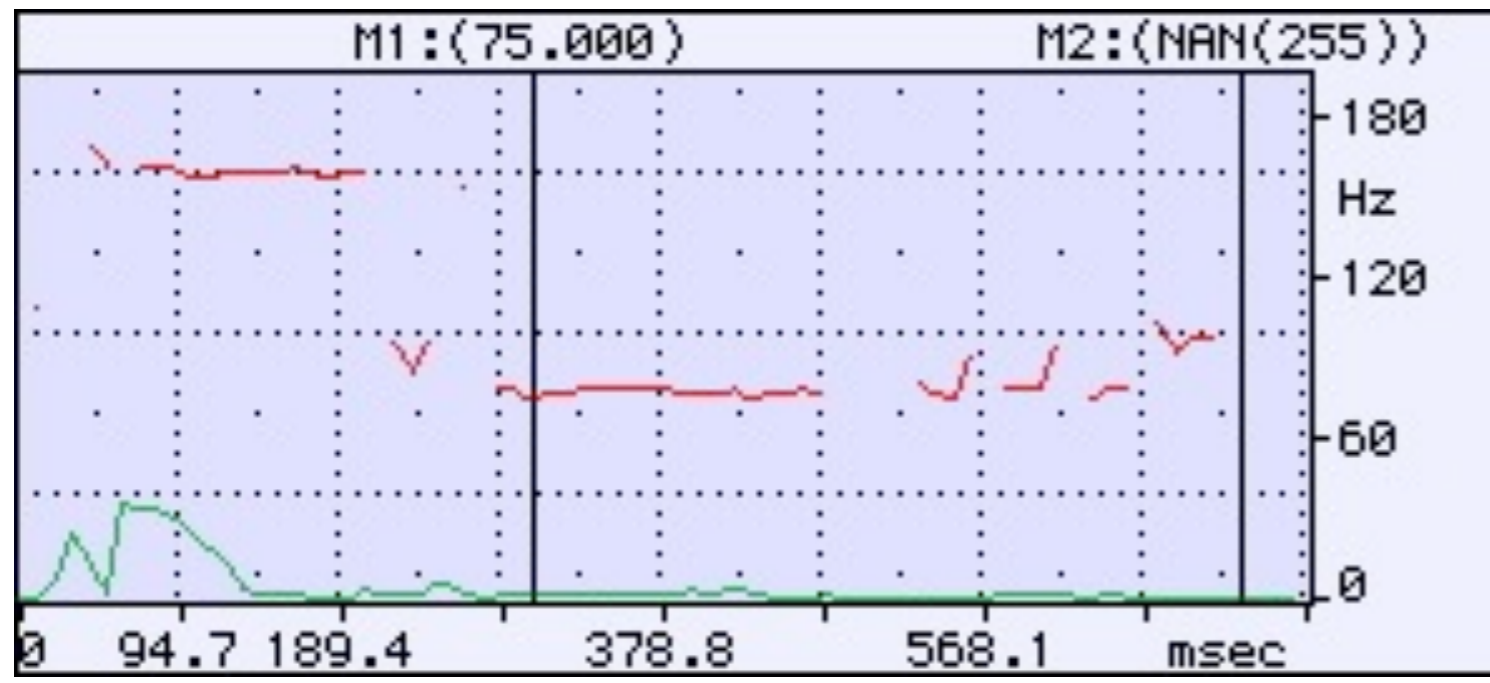


F0 vs. intenzitás

Ő látta?

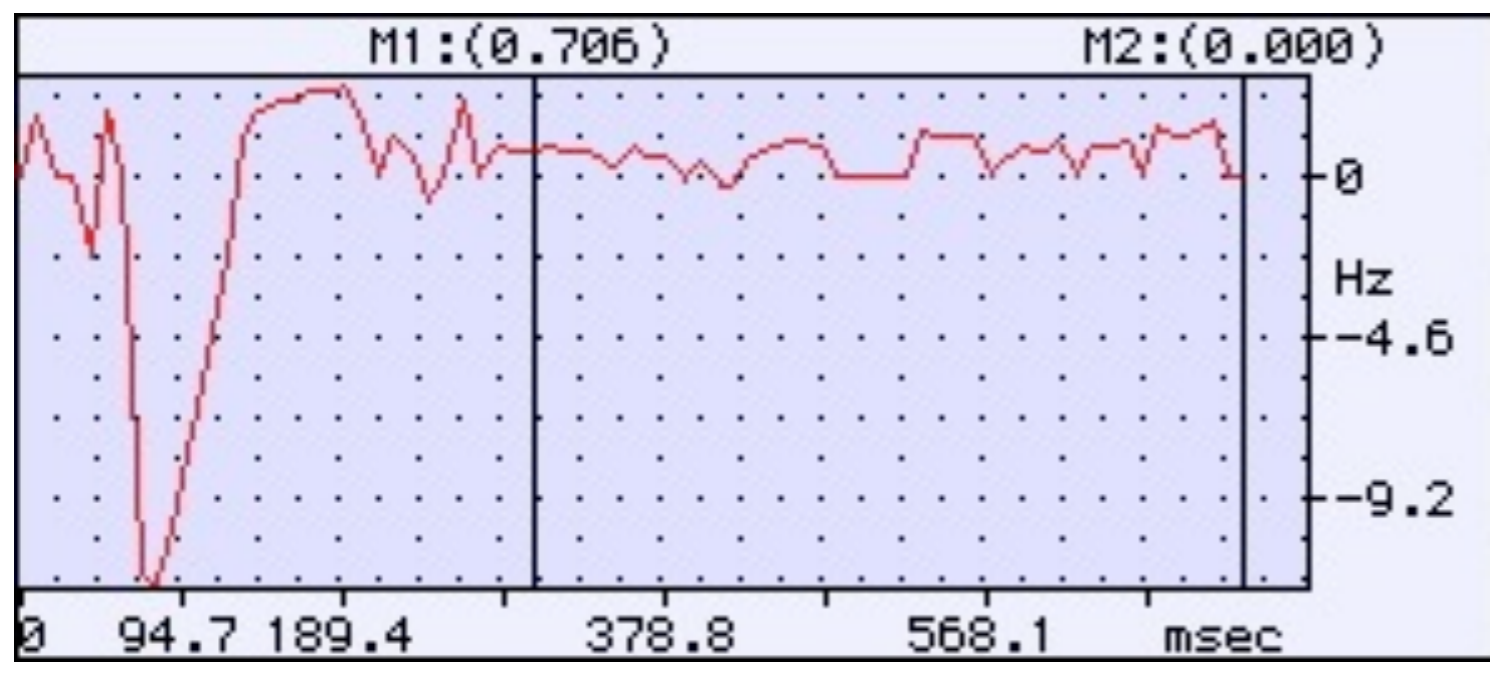


PET

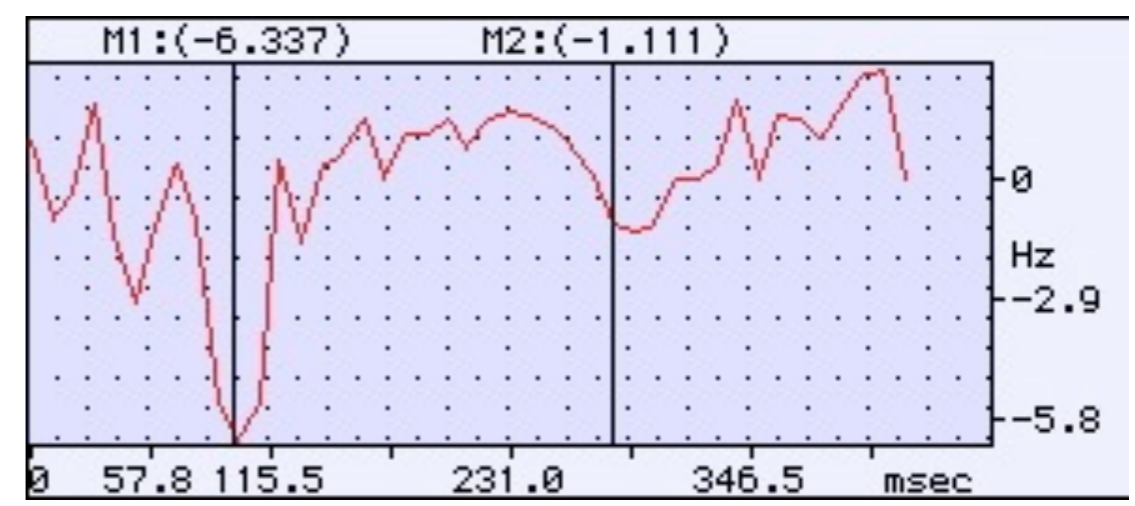


F0 vs. intenzitás

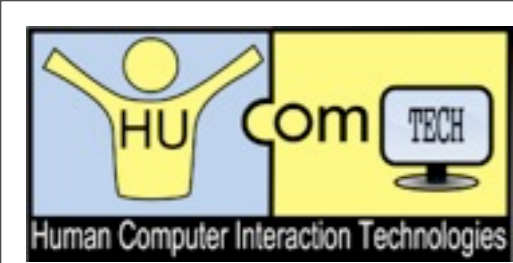
Kati mondja.



PET



PET
mondja



Köszönöm!

<http://hucomtech.unideb.hu/hucomtech>