

A multimodális pragmatikai annotáció jelentősége a számítógépes nyelvészetben¹

Bódog Alexa¹, Abuczki Ágnes¹, Németh T. Enikő²

¹Debreceni Egyetem, Általános és Alkalmazott Nyelvészeti Tanszék
Egyetem tér 1.
4032 Debrecen

²Szegedi Tudományegyetem, Általános Nyelvészeti Tanszék
Egyetem utca 2.
6722 Szeged

{alex.a.weirdling, abuczki.agnes}@gmail.com,
nemethen@hung.u-szeged.hu

Kivonat: Jelen tanulmány egy olyan pragmatikai annotációs eljárást mutat be annak szintjeivel, technikai eszközeivel és kezdeti eredményeivel együtt, amely segítségével lehetővé válik a társalgás bizonyos mozzanatainak automatikus fölismerése és kinyerése, valamint a társalgás szerkezetével és menetével kapcsolatos predikciók megtétele. Az annotációs eljárást a multimodális, spontán hétköznapi társalgásokat tartalmazó magyar HuComTech-korpuszon fejlesztettük ki. Az annotációs rendszer nyelvfüggetlen, univerzális kategóriákkal dolgozik, típusos szerkezetű, az egyes szintek egymásra épülnek benne. Az annotációs szintek az alábbiak: a kommunikatív aktusok szintje, a támogató aktusok szintje, a tematikus kontroll szintje, valamint az adott-új információ szintje. Az eljárás megfelel a jelenleg is kidolgozás alatt álló nemzetközi standardizációs elvárásoknak, követelményeknek.

1 Bevezetés: pragmatika és számítógépes nyelvészet

A számítógépes nyelvészet területén a pragmatikai kutatások és fejlesztések évről-évre nagyobb teret nyernek. E tendencia mögött elsődlegesen az ember–gép interakció (a továbbiakban HCI – human–computer interaction) sikeresebbé tételének motivációja húzódik meg, másodsorban pedig a nyelvtudomány azon komputációs igénye, melynek célja a grammatikai és a pragmatikai kompetencia formális modelljeinek létrehozása [1], [2]. A HCI-alkalmazások sikerének egyik kulcsa azért keresendő a pragmatikában, mert e terület az emberek között zajló kommunikatív nyelvhasználat mibenlétével foglalkozik [3]. A nyelvet különböző kontextusokban, különböző célok

¹ A jelen tanulmány alapjául szolgáló kutatásban Bódog Alexát és Abuczki Ágneszt *A felsőoktatás minőségének javítása a kutatás-fejlesztés-innováció-oktatás fejlesztésén keresztül a Debreceni Egyetemen* című, TÁMOP-4.2.1/B-09/1/KONV-2010-0007 projektazonosítójú program, Németh T. Enikőt pedig az MTA-DE-PTE-SZTE. Elméleti Nyelvészeti Kutatócsoportja támogatta.

elérésének érdekében használjuk, s ennek a bázisnak tükröződnie kell az ember és az általa használt gép kommunikációjában is. E mozzanat fontosságát jelzi az is, hogy a hétköznapi felhasználóknak a gépekhez fűződő viszonya sajátosan későmodern természetű: egyrésztől igényeljük és talán el is várjuk azt, hogy a gépek megkönnyítsék mindennapi életünket (így ebből a szempontból általában pragmatisták vagyunk és a gépeket értéksemleges eszközöknek tekintjük), másrésztől viszont úgy gondoljuk, hogy életünket és céljainkat a gépek igenis befolyásolják, sőt, bizonyos esetekben meg is változtatják – például a gépek használatát a legtöbb esetben explicit módon tanulni kell (ebből a szempontból eszközeinket értékterheltnek tekintjük).³

A fentiek alapján elmondható tehát, hogy a komputációs pragmatika fő kutatási problémái azon jellemzők föltárása és gépi kezelése köré csoportosulnak, melyek a hétköznapi, valós nyelvhasználatot alapvetően meghatározzák. Ennek megfelelően központi helyet foglalnak el a különböző referenciális elemek visszakeresésével kapcsolatos problémák, a nyelvészeti pragmatikából ismert beszédaktusok automatikus generálásának és interpretálásának nehézségei, a beszédaktusokon túlmenően a teljes diskurzusok szerkezete automatikus generálásának és interpretálásának kérdései, valamint az abdukción [2]. E problémák komputációs pragmatikai megoldásai a HCI több kutatási és alkalmazási területén alkalmazhatók, így például dialógusrendszerekben, racionális döntési rendszerekben, vagy akár spontán beszéd felismerő rendszerekben is [1], [4].

A jelenleginél természetesebb HCI megteremtését célzó projektek között megkülönböztetett fontossággal bírnak a kommunikáció multimodalitását alapul vevő kutatások, melyek során nemcsak a beszélt nyelvi kommunikatív információmanipuláció jellemzőit tárjuk föl, hanem figyelembe vesszük a nem verbális akusztikus, valamint a vizuális tartományból érkező információkat is.

A fent említett problémák megoldásához többféle adatgyűjtési és elméletalkotási modellt hívhatunk segítségül, melyek közül a legelterjedtebb módszer a korpusz- és adatbázis-építésen alapuló adatkinyerés, illetve elméleti általánosítások megtevése.

Számítógépes nyelvészek és informatikusok számos sémát fejlesztettek ki azon törekvés során, hogy standardizált kódnyelvet és terminológiát hozzanak létre különböző korpuszannotációk számára. Mivel a korpusz- és adatbázis-építés fő kritériuma a standardok követése és az interoperabilitás, ezért ezeket a sémákat általában XML-ben kódolják, amely lehetővé teszi a gépi feldolgozást. A nyelv verbális aspektusainak kódolása mellett szintén standardizált rendszerré fejlődött a nem verbális jegyek kódolása is, melyekre sajátos kódnyelvek születtek, mint például a nemzetközi élvonalbeli, arcizommozdításokat figyelembe vevő Ekman-féle FACS-kódrendszer (Facial Action Coding System).⁴ A multimodális kódolósémák közül úttörőként emelkedett ki a MUMIN⁵ multimodális kódrendszer a gesztusok és arckifejezések személyközi kommunikációban betöltött szerepének tanulmányozására. A fenti sémákhoz hasonlóan a HuComTech kutatócsoport is egy többszintű, multimodális an-

2 Vigyázat, nem pragmatikusok, csak pragmatisták!

3 A gépekhez fűződő attitűdjeinkről jó áttekintést ad Ropolyi László [5].

4 A FACS manuáljának részlete elérhető az alábbi weboldalon: <http://face-and-emotion.com/dataface/facs/manual/TitlePage.html>

5 MUMIN: <http://www.ling.helsinki.fi/kit/2006k/clt310mmod/MUMIN-coding-scheme-V3.3.pdf>

notációs rendszert épített ki, amely figyelembe veszi a kommunikáció verbális akusztikus, nem verbális akusztikus és vizuális jellemzőit is, így különféle multimodális természetű lekérdezésekre és modellépítésre is alkalmas.

Ugyanakkor nem szabad elfelejtkeznünk arról, hogy a korpusz és adatbázis használata egy elméleti döntés, ahogyan az is, hogy milyen annotációt készítünk, szintaktikait, morfológiáit vagy pragmatikait, továbbá, hogy a választott típusú annotációs rendszerünk milyen alapegységekkel és szintekkel dolgozik. A HuComTech-korpuszon alkalmazott multimodális pragmatikai annotáció mögött az az elméleti megfontolás húzódik, hogy a kommunikáció során a kommunikációs partnerek egyszerre, szimultán módon veszik figyelembe a különböző elérhető modalitásokból származó stimulusokat. Ezen elméleti döntés értelmében válhatott a multimodális pragmatikai annotáció alapegységévé a kommunikatív aktus.

Jelen tanulmány a kommunikatív aktusok generálására és interpretációjára összpontosít a HuComTech-korpusz vizsgálata és multimodális pragmatikai annotációja alapján. Célunk kettős: egyrésztől szeretnénk bemutatni egy olyan, saját fejlesztésű multimodális pragmatikai annotációs rendszert, mely segítségével oly módon tudjuk leírni és értelmezni a személyközi kommunikatív viselkedéseket, hogy az tevékenyen hozzájárulhasson a beszélt ember-gép interakciót lehetővé tévő dialógusrendszerek modellálásához és kivitelezéséhez. Közvetett célunk pedig az, hogy rávilágítsunk arra, hogy a hagyományosan nem formális természetű nyelvészeti pragmatika aktívan képes hozzájárulni a számítógépes nyelvészethez (és viszont), valamint hogy ez a hozzájárulás nem öncélú. Fontos kiemelni azt, hogy kutatásunk e tanulmány elkészítésekor még nem zárult le – az annotáció jelenleg is folyik, így végleges elméleti általánosítások levonására, valamint eredményeink dialógusrendszerbe történő integrálására egyelőre még nem volt módunk. Ennek ellenére annotációs rendszerünk előnyei már most kézzelfoghatók.

Céljainknak megfelelően elsőként röviden bemutatjuk a HuComTech-csoport által épített korpuszt, annotálásunk terepét, majd pedig a QANNOT-annotációs eszközt. Előadásunk legfontosabb részében saját multimodális pragmatikai annotációs rendszerünk szintjeit mutatjuk be példák segítségével, valamint az annotálás eszközét, az annotációs folyamatot és további kutatási terveinket. Zárásként kísérletet teszünk tanulmányunk metaelméleti reflexiójára is, hogy kimutassuk a nyelvészeti pragmatika és a számítógépes nyelvészet egymásra gyakorolt hatását.

2 A HuComTech-korpusz multimodális pragmatikai annotálásának elméleti alapjai

Multimodális pragmatikai annotációs rendszerünk alapjait egy korábbi tanulmányunkban részletesen kifejtettük [6]. Jelen tanulmányban céljainknak megfelelően arra összpontosítunk, hogy rámutassunk a hagyományos nyelvészeti pragmatika és a számítógépes nyelvészet közös metszéspontjaira, így annotációs rendszerünk elméleti alapjait is e nézőpontból mutatjuk be.

A pragmatikai annotáció a társalgás szegmentálását és címkézését jelenti, melynek során nyelvi információt adunk hozzá a nyelvi szegmensekhez, valamint a nem verbá-

lis kommunikatív viselkedést is szegmentáljuk és címkézzük. A pragmatikai annotáció elsősorban a beszélő szándékának megfelelő, és nem csupán a formában (a felszíni szerkezetben) tükröződő kommunikatív funkciókat jelöli meg, hiszen a sikeres kommunikáció feltétele az, hogy a hallgató/címzett ugyanúgy értelmezze a beszélő/feladó megnyilatkozását és szándékait, ahogyan ő (a beszélő) is kívánta [6].

Multimodális pragmatikai annotációs rendszerünk alapját a *kommunikatív aktusok* képezik. A kommunikatív nyelvhasználat e minimális alapegységei nyelvi szempontból megnyilatkozások [7], amelyek társalgási fordulókba, a fordulók szomszédsági párokba, a párok pedig koherens diskurzusokba szerveződnek. A beszélt nyelvi dialógusokat a társalgáselemzésben általában fordulóokra szokás szegmentálni, ám mivel ezek a szegmensek nagyon hosszúak is lehetnek, ezért előnyösebb őket további funkcionális egységekre, kommunikatív aktusokra tagolni. A kommunikáció során minden szint sajátos elvek és megszorítások alapján szerveződik. A nyelvészeti pragmatika oldaláról nézve a kommunikatív aktusok multimodális illokúciós aktusok. Ilokúciós aktusok, mivel a bennük kifejezett beszélői és szándékolt hallgatói attitűdök alapján szerveződnek, így előtérbe kerülnek a kommunikációban jelen levő intenciók, s multimodálisak, mivel a verbális közlés mellett figyelembe vesszük a vizuális (a gesztusokkal, valamint a különböző arckifejezésekkel támogatott) és a nem verbális akusztikus (prozódiai) információkat is. Az illokúciós aktusok nyelvészeti pragmatikai kutatásai rámutatnak arra, hogy a partikuláris illokúciós aktusok száma igen magas, így ezek vizsgálata parttalaná válhat mind a kategorizáció, mind a csoportosítás tekintetében. Például a kérésnek mint illokúciós aktus fajtának rengeteg „alfaja” különböztethető meg (kérés, parancs, könyörgés, utasítás, kíváncsi stb.), s ezek az aktusok ráadásul még nyelvfüggő természetűek is (az egyik nyelvben megvannak, a másiktól pedig hiányoznak). Amennyiben magas szinten általánosító modellt kívánunk létrehozni, úgy ki kell küszöbölnünk a nyelvfüggő, partikuláris kategóriákat – túl kell lépni az „egy jelenség = egy szabály” típusú leírásokon. Multimodális pragmatikai annotációs rendszerünkben ezt a problémát úgy oldottuk meg, hogy nem partikuláris aktusokat, hanem aktustípusokat különböztettünk meg egymástól a Bach és Harnish által kidolgozott illokúciós aktustipológia alapján [8]. A típusos megközelítés egyik pozitívuma az tehát, hogy valamilyen szempont alapján (jelen esetben az aktusban kifejezett beszélői és a szándékolt hallgatói attitűdök alapján) osztályokba, típusokba sorolja a példányszintű (token) jelenségeket, így a rendszer alkalmas lesz általános szabályszerűségek megállapítására, s ebből eredően predikciók megtételére.

Rendszerünkben például a kérések, parancsok, kíváncsi stb. egységesen a direktív aktusok típusába tartoznak. A direktív aktusok olyan aktusokat tartalmaznak, melyek propozicionális tartalma a hallgató egy elvárt/preferált jövőbeli cselekedetére vonatkozik, s amelyek kifejezik a beszélő azon szándékát, hogy a hallgató a szóban forgó aktus hatására hajtsa végre a jövőbeli cselekedetet [6]. A direktívek mellett megkülönböztettünk konstatívokat (melyek a beszélőnek egy propozicionális tartalomhoz fűződő hiedelmét fejezik ki úgy, hogy a beszélő mindeközben szándékozza azt is, hogy az aktus propozicionális tartalmát feldolgozza és higgye a hallgató is), kommisszívokat (amelyek a beszélő azon szándékát fejezik ki, amellyel elkötelezi magát egy jövőbeli aktus megtételére) és ún. viselkedő aktusokat is (*acknowledgement*, a beszélő valamilyen affektív, érzelmi, attitűdbeli viszonyulását fejezik ki a hallgató felé). A társalgásban előfordulnak olyan esetek is, amikor a megnyilatkozás semmifé-

le propozicionális tartalommal nem rendelkezik, s a megnyilatkozás konkrét illokúciós ereje nem azonosítható.⁶ Ebben az esetben a none (nem azonosítható) címkét alkalmazzuk az annotáció során.

A típusos megközelítés másik előnye az univerzalitás: míg a partikuláris aktusok nyelvfüggőek, addig az aktusok típusai nagy valószínűséggel nyelvfüggetlenek [9]. Az univerzális jelenségek mögött meghúzódó szabályszerűségek föltárása a nyelvészeti pragmatikában és a számítógépes nyelvészetben egyaránt fontos: a pragmatika számára azért, mert absztrakt, általános érvényű megállapításokat tudunk tenni a nyelvhasználatra vonatkozóan, a számítógépes nyelvészet számára pedig azért, mert e megállapításokat fölhasználva túl tud lépni a statisztikai alapú alkalmazásokon.

A kommunikatív aktusok mellett az úgynevezett támogató aktusokat is annotáljuk a multimodális pragmatikai annotáció során. Ezek az aktusok nem bírnak önálló illokúciós értékkel, ehelyett kiegészítik, támogatják a velük egy fordulóban szereplő kommunikatív aktust. Ezen aktusok annotálása azért fontos a nyelvészeti pragmatika számára, mert segítségükkel számot tudunk adni egyrészt az “interakcióban levés” mozzanatairól, másrészt a társalgás formai jegyek alapján történő szegmentálásáról. E két mozzanat a számítógépes nyelvészet számára is fontos: az interakcióban való részvételnek pragmatikai funkciójú multimodális jelölői vannak, például a visszajelzés (backchannel), mely történhet bólogatással, hümmögéssel, ühümözéssel stb. Emellett a társalgásban olyan formai jelölők, például diskurzusjelölők és udvariassági markerek is részt vesznek, melyek segítségével könnyen azonosíthatóvá válnak a megnyilatkozásokat alkotó kommunikatív aktusok típusai. Például hiába hangzik el egy kérdő intonációjú megnyilatkozás, ha a végén szerepel a *légy szíves* kifejezés vagy a *kérlek* szócska: tudjuk, hogy a megnyilatkozás ebben az esetben kérés lesz.⁷ Multimodális pragmatikai annotációs rendszerünkben a támogató aktusok közül a visszajelzéseket, az udvariassági markereket, valamint a javításokat (melyek során a beszélő a saját partikuláris kommunikatív aktusához fűződő attitűdjét változtatja meg) jelöljük. Távolati terveink között szerepel a diskurzusjelölők annotálása is.

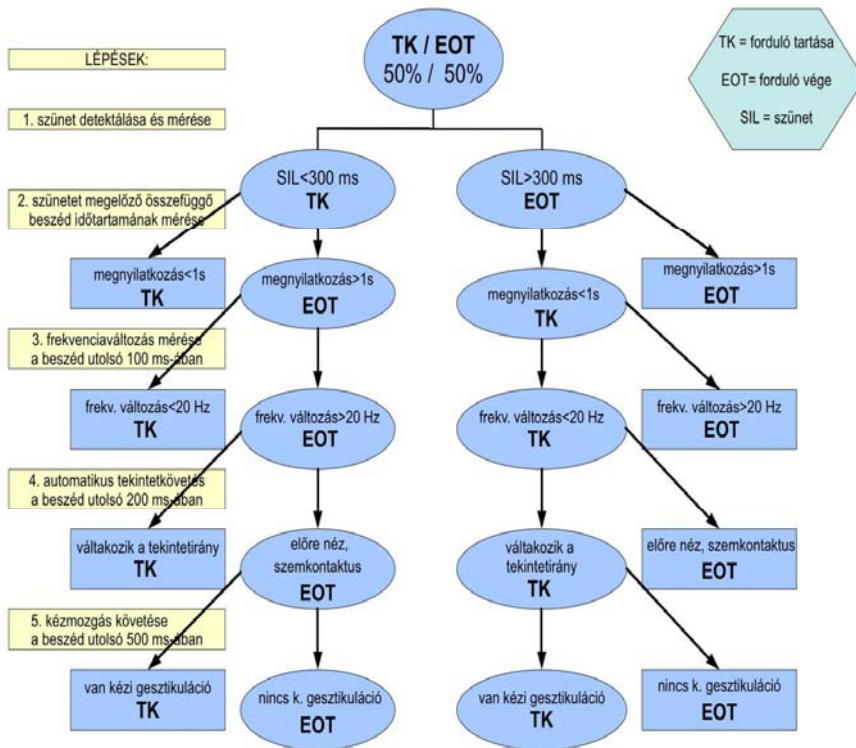
Ahogy korábban említettük, a kommunikatív aktusok és a támogató aktusok együttesen megnyilatkozásokba, a megnyilatkozások pedig társalgási fordulókbá szerveződnek. A fordulók tehát több megnyilatkozást, s azon belül több kommunikatív aktust is tartalmazhatnak, határaitak a beszélőváltás jelöli ki. A beszélőváltás azonban nem véletlenszerűen történik a társalgásban: például egy kérdés elvárt következménye a válasz, egy javaslat elvárt következménye az elfogadás, a nyugtázás. Az egymásra következő fordulókból olyan szomszédsági párok bontakoznak ki, ahol a pár második tagja rendszerint az első párra adott elvárható, preferált válasz. Tehát az interakció elméleti modellezése során szintén érdemes párszekvenciákba összekapcsolni a kommunikatív aktusokat. Dialógus-modellekben általában két kommunikatív aktus alkot egy párszekvenciát [10]: az első kommunikatív aktust a gép nyújtja, a második aktus pedig az (emberi) felhasználó fordulója. Mivel a kommunikatív aktusok jellemzően előre megjósolható sorrendben követik egymást (pl. kérdés-válasz és kérés-teljesítés szekvenciákban) [11], [12], így az egyes aktusok jellemző jegyeinek

⁶ Ilyen eset például a *Jaj!* fölkiáltás.

⁷ Nyelvészeti pragmatikai terminussal élve a konvencionálisan indirekt illokúciós aktusok automatikus felismerésének lehetőségeit kívánjuk fölteni.

az annotációból való kinyerése megkönnyítheti a dialógusrendszer betanítását azok felismerésére és megfelelő válaszok automatikus generálására is. Ha adott az egyik rész, előre jelezhető a másik [1].

Annotációs rendszerünk lehetővé teszi azt, hogy a többi HuComTech-annotációval egybefűzve megvizsgáljuk a társalgási fordulók szomszédsági párokba való szerveződésének mozzanatait is. E vizsgálat pedig elvezethet minket kommunikatív aktusok közötti döntéshozást segítő, következő aktust jósló döntési fák létrehozásához is. Mivel a kommunikatív aktusok automatikus felismerésének, predikciójának és generálásának első lépése és egyben alapfeltétele a beszélőváltás predikciója is, ezért Abuczki Ágnes [13] kvantitatív vizsgálatokkal, adatbázis-lekérdezésekkel a fordulólezáras és a lehetséges váltási pont tipikus jegyhalmazát gyűjtötte össze Troung és munkatársai [14] modelljéből kiindulva, majd ezeket a jellemzőket vizuális jegyekkel kiegészítve egy döntésfába rendezte (1. 1. ábra).



1. ábra: Döntésfá a forduló lezárásának ('end-of-turn', rövidítése: EOT) és a forduló tartásának ('turn-keep', rövidítése: TK) megkülönböztetésére multimodális jegyek alapján [13].

Az 1. ábrán látható döntésfá a fordulózárás ('end-of-turn', rövidítése: EOT) és a forduló tartásának ('turn-keep', rövidítése: TK) megkülönböztetésére vállalkozik. A döntésfán látható öt lépés közül az első három akusztikai tényezőket, az utolsó két lépés pedig vizuális tényezőket tartalmaz. A beszélőváltás predikciójával egyidejűleg a szomszédsági párok tipikus mintázatának megfelelően, az egyes kommunikatív aktustípusok lekérdezések után kapott megkülönböztető jegyeire támaszkodva, a jegyeket a fenti példához hasonlóan döntési fába rendezve a következő kommunikatív aktust megjósoló modellt hozhatunk létre. A pragmatikai annotáció mellett az audio- és videoszinten is annotált HuComTech-korpusz megbízható kiinduló bázisa lehet az egyes kommunikatív aktusok együtt járó jellemzői összegyűjtésének, ami hozzájárulhat az emberi beszélő által végrehajtott aktusok automatikus felismeréséhez. Az egyelőre még csak vázlatosan modellált dialógusrendszer feladata elsősorban „csupán” a fordulók végének detektálása lesz a tipikus fordulóvégi jellemzők (audio- és vizuális markerek) együttes előfordulása és bizonyos időtartamú események egymást követő sorrendje alapján. A megnyilatkozás végének detektálása után pedig a gépi ágens felteheti a beépített forgatókönyvének (*scenario*) megfelelő következő kérdést. Így a kérdések és válaszok láncából felépül a dialógus. A szomszédsági párok sorozatából épül ki a társalgás egésze, melynek során akár több témát is egymásba fűzhetünk. Ezért annotációnkba a tematikus kontroll szintjét is bevezettük, mellyel célunk az volt, hogy korrelációkat tudjunk megállapítani az egyes kommunikatív aktusok szekvenciális szerveződése, a fordulókezelés, valamint a globális diskurzusszerveződés mozzanatai között. Annotációs rendszerünkben megkülönböztetjük a témakidolgozás, az egyes társalgási témák motivált egymásba fűzése, illetve a motiválatlan témaváltás mozzanatait.

A pragmatikai annotáció utolsó szintjén a társalgás univerzumába kerülő új lexikai információkat jelöltük. Erre azért volt szükség, hogy a későbbiekben megvizsgálhassuk azon hipotézisünket, amely szerint az új információ bevezetése élénkebb, erőteljesebb gesztikulációval és nagyobb intenzitással jár együtt. [13] kvalitatív elővizsgálata a szemantikailag új lexikális információ kézi bejelölése után azt az eredményt hozta, hogy a gesztus csúcspontja (ún. *stroke*) és a szemantikailag legfontosabb verbális egység gyakran egybeesik. Ezt a feltételezést kvantitatív módszerekkel, vagyis a tervezett lekérdezések statisztikai elemzésével is kívánjuk igazolni a HuComTech-korpuszban.

3 A multimodális pragmatikai annotációs séma

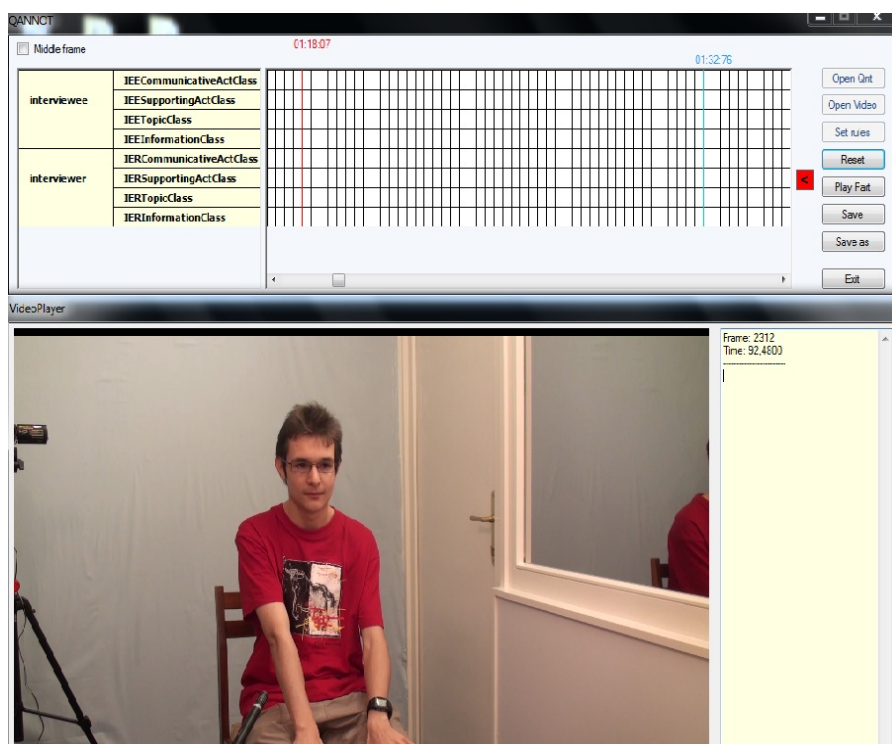
A HuComTech-projekt multimodális pragmatikai annotációjának sémáját az alábbi táblázatban összegezzük:

1. táblázat: A HuComTech-projekt multimodális pragmatikai annotációs sémája.

kommunikatív aktusok típusai (the level of communicative act types):
konstatívok (constatives) = ítélkezők: válaszadás, megerősítés, informálás, predikció, visszaemlékezés
direktívák (directives) = végrehajtók: kérés, parancs, javaslattétel
kommisszívok (commissives)= elkötelezők: beleegyezés (pl. egy fogadásba), följánlás, ígéret
viselkedők (acknowledgements): üdvözlés, bűcsúzás, elfogadás (pl. meghívásé)
indirekt (indirect)
nem azonosítható (none)
támogató aktusok szintje (the level of supporting acts):
visszajelzés (backchannel)
udvariassági marker (politeness marker)
javítás (repair)
nincs aktus (none)
tematikus kontroll szintje (the level of thematic control):
témakezdeményezés (topic initiation)
témakidolgozás (topic elaboration)
témaváltás (topic change)
információ szintje (the level of information type):
adott (given)
új (new)

A multimodális pragmatikai annotáció eszközt, a QANNOT-programot Szeghalmy Szilvia (Debreceni Egyetem) hozta létre 2010-ben a HuComTech-csoport számára. A QANNOT-ban az annotáció egysége – amely egységekhez *timestamp*eket (kezdő- és végpontokat) lehet rendelni – a *frame*. A pragmatikai annotáció jelenleg beállított szegmentálási egysége 8 frame per second, vagyis nyolc frame reprezentál egy másodpercet. Az annotáció során az annotátorok elsőként a kommunikatív aktusok típusainak címkéit helyezik el az annotálni kívánt videó idővonalán. Ezután a támogató aktusok címkézése következik, majd a témaváltás. Végezetül az adott-új információ címkézése történik. Mivel a QANNOT egyszerre jeleníti meg az összes annotációs szintet, így az annotátorok szimultán módon össze tudják hasonlítani és szinkronizálni egymással a különböző szinteken elhelyezett címkéket.⁸

⁸ Ez nemcsak a multimodális pragmatikai annotáción *belül* fontos, hanem akkor is, amikor a különböző annotációkat egybe kívánjuk vetni, s korrelációkat megállapítani például a Praatban zajló prozódiai és a QANNOT-ban zajló multimodális pragmatikai annotáció címkéi között.



2. ábra: A multimodális annotáció felhasználói felülete a QANNOT-programban.

4. Tervezett lekérdezések a HuComTech-korpuszon

A kutatás következő szakaszában (a 2011-es MSzNy konferencia időpontjáig) kvantitatív elemzést kívánunk végezni adatbázis alapú címkelekérdezések segítségével, melyekről előadásunkban részletesen be fogunk számolni. Multidimenziós vizsgálatot fogunk végezni, vagyis a dialógusok horizontális és vertikális szerkezetét egyaránt elemezni fogjuk a különböző típusú (audio, video, szintaktikai és pragmatikai) annotációk bizonyos szintjeinek (a diskurzus, a tekintetirány, a kommunikatív és támogató aktusok, valamint a tematikus kontroll szintjének) szimultán többszintű megjelenítése és együttes előfordulásukra vonatkozó címkelekérdezések segítségével.

A horizontális (szekvenciális) elemzés részeként az annotáció horizontális vetületét fogjuk vizsgálni, vagyis ennek segítségével az időben egymást követő jelenségek (elsősorban kommunikatív aktusok) mintázatát próbáljuk feltárni.

A vertikális címkeelemzés keretében pedig audio-, video- és pragmatikai címkék együttjárását keressük: első lekérdezéseink során azt vizsgáljuk meg, hogy az egyes kommunikatívaktus-típusok (konstatív, direktív, kommisszív, viselkedő, indirekt) jellemzően milyen embléma típusú gesztusokkal (figyelem, egyetértés, nem egyetértés,

viSSzautasítás, kételkedés, számok és alak, valamint méret mutatása⁹) és milyen arcki-fejezésekkel (semleges, boldog, meglepett, szomorú, elgondolkodó, feszült¹⁰) (a kategóriák részletes bemutatásáért l. [15]) járnak vagy kezdődnek együtt (vagyis melyik kommunikatív aktusba esik bele egy gesztus vagy arckifejezés kezdőpontja). Ezeket az eredményeket olyan formában szeretnénk megkapni, hogy hány-hány darab emb-lématípus jelenik meg az egyes kommunikatív aktus-típusok végrehajtása közben. Vagyis a fenti vertikális természetű lekérdezések fő célja az egyes aktus-típusokat kísérő nem verbális-vizuális, nem verbális-akusztikus és verbális jegyek felfedése, amelyek szisztematikus rendszerbe foglalása és explicitté tétele elvezethet minket a kommunikatív aktusok automatikus felismeréséhez.

Következő lépésként, a szekvenciális (horizontális) elemzés során a kommunikatív aktusok egymást követő tipikus sorrendjeit szeretnénk megállapítani. Ezzel validálni szeretnénk a szomszédsági párok [11] által felállított tipikus láncolat alkalmazhatóságát magyar spontánbeszédkorpuszon is. Ezt a lekérdezést úgy fogjuk elvégezni, hogy diskurzusszinten¹¹ a záró (turn give közben végrehajtott) és a nyitó (turn take közben végrehajtott) kommunikatív aktusokat párosítjuk, majd a kapott aktuspárokat csoportosítjuk és megszámloljuk. Mivel a párszekvencia első fele előrejelzi a második felét - különösen formális, kanonikus szituációkban -, így ez a megközelítés grafikus és multimodális felhasználói felületek működtetéséhez egyaránt megfelelő feltételeket biztosít. Eredményeinkkel ezáltal nemcsak a kommunikatívaktus-típusok felismeréséhez, hanem azok automatikus generálásához és összefonásához, diskurzusba kapcsolásához is célunk hozzájárulni.

Következő lekérdezésünk arra a kérdésre keresi a választ, hogy az egyik beszélő által végrehajtott visszajelzés (*backchannel*) a másik beszélő által végrehajtott mely kommunikatívaktus-típusba és hány alkalommal esik bele.¹² Ezzel azt kívánjuk feltárni, hogy leggyakrabban milyen aktustípusra következik reakcióként a visszajelzés, vagyis mi a visszajelzés leggyakoribb funkciója.

A kommunikatív aktusok akusztikai markereinek feltárásához a Praat-program [16] áll rendelkezésünkre. A Praat-programban – melyben a HuComTech-korpusz audioannotációja zajlik - a spektrogram horizontális irányban mutatja az időtartamot, vertikális irányban pedig a frekvencia (hangmagasság) skálázását (Hz mértékegységben). A fenti adatok millisecundumonkénti értékeinek feltöltése lehetővé teszi a felvételek fonetikai elemzését és fonetikai jellegű (például intenzitásra és alaphangfrekvenci-

9 Sémánkban a címkék angolul szerepelnek: attention, agree, disagree, refusal, doubt, numbers, size.

10 Sémánkban a címkék angolul szerepelnek: natural, happy, surprised, sad, recalling, tensed.

11 A HuComTech-korpusz audioannotációja tartalmaz egy diskurzusszintet, ahol a társalgás fordulókra van bontva [13]. A fordulókat a következő címkék jelölik: T (turn taking: a forduló átvétele/kezdet), K (turn keeping: 'a forduló megtartása'), G (turn giving: 'forduló átadása') és BC (backchannel: 'a hallgató fél rövid, figyelmet jelző visszajelzése'). Egy beszélő fordulóján belül akár több kommunikatív aktus is előfordulhat, tehát az audioannotáció további információkkal bővíti a pragmatikai szinten.

12 Olyan visszajelzéseket (BC) vizsgálunk, amelyek kezdőpontja belesik a másik beszélő által végrehajtott kommunikatív aktus időtartamába. Aktustípusonként egyesével szükséges lekérdezni a kommunikatív aktusok darabszámát és időtartamát, illetve a bennük végrehajtott visszajelzések darabszámát.

ára vonatkozó) lekérdezések végrehajtását. Ezek után elsődleges célunk az egyes kommunikatívaktus-típusok átlagos intenzitásminimumának, -maximumának és -átlagának lekérdezése lesz, annak érdekében, hogy ezekkel az eredményekkel is hozzájáruljunk az egyes aktustípusok megragadásához és formalizált leírásához, amely a későbbiekben elvezethet minket a beszélő kommunikatív szándékának automatikus felismeréséhez, illetve előrejelzéséhez.

5 Összegzés

A jelen tanulmányban bemutatott pragmatikai annotációs rendszer fő előnye abban rejlik, hogy univerzális kategóriákkal dolgozik, vagyis a felvételek nyelvtől függetlenül univerzálisan alkalmazható, hiszen a kommunikatív és a támogató aktusok típusai, valamint a tematikus kontroll tulajdonságai egyaránt univerzális jellemzői a társalgásnak. A rendszer interoperábilis XML-sémája lehetővé teszi az annotációs szempontok, annotálandó kommunikatív jelenségek bővítését újabb szintek és címkék bevezetésével. Ugyanakkor a fölöslegessé vált szintek és címkék is törölhetők (például a *none* címkét bevezetését követően hamarosan töröltük). A fordulók mint strukturális elemek és a kommunikatív aktusok típusai mint funkcionális elemek együttes szerepeltetése lehetővé teszi, hogy a fordulókból kibontakozó szomszédsági párokhoz megfelelő kommunikatívaktus-típusokat tudjunk rendelni. Mivel a QANNOT-program képes egyszerre megjeleníteni az összes annotációs szintet, így lehetővé válik az egyes szintek címkéinek szimultán összehasonlítása (például a kommunikatív aktusok összevetése az audio- és videoannotáció címkéivel), illetve a címkestatisztikai adatbázisba való feltöltés után bizonyos kommunikatív jelenségekre jellemző multimodális jegyhalmazok explicit formában történő felfedése. Ez közelebb vihet minket olyan multimodális jegyhalmazok meghatározásához és finomításához, amelyek segítségével nagy biztonsággal meg tudjuk jósolni a következő forduló kommunikatív aktusát/aktusait a társalgásban.

Mindezen megfontolásokat figyelembe véve a HuComTech-korpusz pragmatikai annotációja tevékenyen hozzájárul az ember-gép kommunikációs technológiák nyelvészeti aspektusainak modellezési lehetőségeihez. Ha a számítógépes nyelvészeti adatbázisokra alapozva kívánja a kommunikációt modellálni, akkor annak szüksége van a pragmatikára, hiszen jól megalapozott pragmatikaelméleti döntéseket igényel annak meghatározása, hogy milyen legyen a társalgások pragmatikai annotációja. Ugyanakkor a pragmatika számára is nyereséggel jár a számítógépes nézőpont, mert rákényszeríti a pragmatikusokat, hogy a kommunikatív nyelvhasználatra vonatkozó megállapításait explicit formában fogalmazzák meg, úgy, hogy azok formalizálásra alkalmasak legyenek és ezáltal algoritmizálhatóakká váljanak.

Bibliográfia

1. Bunt, H., Black, W.: The ABC of computational pragmatics. In: Bunt, H., Black, W. (eds.): *Abduction, belief and context dialogue: Studies in computational pragmatics*. John Benjamins, Amsterdam (2000) 1–46
2. Jurafsky, D.: Pragmatics and computational linguistics. In: Horn, L. R., Ward, G. (eds.): *The handbook of pragmatics*. Blackwell, Oxford (2002) 578–604
3. Németh T. E.: Pragmatika. In: Kiefer F. (szerk.): *A magyar nyelv*. Akadémiai Kiadó, Budapest (2006) 222–261
4. Jurafsky, D., Bates, R., Coccaro, N., Martin, R., Meteer, M., Ries, K., Shriberg, E., Stolcke, A., Taylor, P., Van Ess-Dykema, C.: Automatic detection of discourse structure for speech recognition and understanding. In: *Proceedings of the 1997 IEEE Workshop on Speech Recognition and Understanding*, Santa Barbara (1997) 88–95
5. Ropolyi L.: *Technika és etika*. In: Fekete L. (szerk.): *Kortárs etika*. Nemzeti Tankönyvkiadó, Budapest (2004) 245–292
6. Abuczki Á., Bódog A., Németh T. E.: A multimodális pragmatikai annotáció elméleti alapjai az ember–gép kommunikáció modellálásában. In: Németh T. E. (szerk.) *Ember–gép kapcsolat. A multimodális ember–gép kommunikáció modellezésének alapjai*. Tinta Könyvkiadó, Budapest (2011, megjelenés alatt)
7. Németh T. E.: Megnyilatkozás: típus - példány. *Néprajz és Nyelvtudomány* Vol. 35 (1994) 69–101
8. Bach, K., Harnish, R. M.: *Linguistic communication and speech acts*. MIT Press, Cambridge (1979)
9. Verschueren, J.: *Understanding pragmatics*. Arnold, London (1999)
10. Bogdan, C., Kaindl, H., Falb, J., Popp, R.: Modeling of interaction design by end users through discourse modeling. In: *Proceedings of the 13th international conference on Intelligent user interfaces*. ACM, New York (2008)
11. Levinson, S. C.: *Pragmatics*. Cambridge University Press, Cambridge (1983)
12. Schlegoff, E. A.: *Sequence organization in interaction: A primer in conversation analysis*. Cambridge University Press, Cambridge (2006)
13. Abuczki, Á.: A multimodális interakció szekvenciális elemzése. In: Németh T. E. (szerk.) *Ember–gép kapcsolat. A multimodális ember–gép kommunikáció modellezésének alapjai*. Tinta Könyvkiadó, Budapest (2011, megjelenés alatt)
14. Troung, K. P., Poppe, R., Heylen, D.: A rule-based backchannel prediction model using pitch and pause information. In: *Proceedings of Interspeech (2010)* 3058–3061
15. Földesi, A.: Unimodális funkcionális annotáció a HuComTech-korpuszban. In: Bódog, A. (szerk.): *Az ember–gép kommunikáció technológiájának elméleti alapjai*. IKUT zárókötet. (előkészületben)
16. Boersma, P., Weenink, D.: *Praat: doing phonetics by computer 5.0.02*. Institute of Phonetic Sciences, University of Amsterdam (2007) <http://www.praat.org>