Az ember–gép kommunikáció elméleti–technológiai modellje és nyelvtechnológiai vonatkozásai

Hunyadi László – Földesi András – Szekrényes István – Staudt Alexandra – Kiss Hermina – Abuczki Ágnes – Bódog Alexa

Debreceni Egyetem, Általános és Alkalmazott Nyelvészeti Tanszék, Debrecen
hunyadi@ling.arts.unideb.hu; andras.foldesi1@gmail.com; xepenator@gmail.com; crysaetos@gmail.com; kissh3@gmail.com; abuczki.agnes@gmail.com; alexab@unideb.hu

A tanulmány a nyelvészet, a kommunikációkutatás, a pszichológia, az informatika és a kognitív robotika találkozásánál elhelyezhető kutatás első eredményeiről kíván számot adni. A kutatás komplexitását a feladatmeghatározás indokolja: szeretnénk megérteni a dialógusokban megjelenő ember–ember kommunikáció alapszerkezetének azon aspektusait, amelyek relevanciával bírnak az ember–gép interakcióban és amelyekről úgy gondolhatjuk, hogy technikailag megvalósíthatók. Ezen fő szempontok vezérlik egy elméleti–technológiai modell fő vonalainak a meghatározását éppúgy, mint egy multimodális korpusz létrehozását. A tanulmányban először e modellt mutatjuk be, majd a korpusz fő moduljait ismertetjük és elemezzük a korpusz alapján létrehozott HuComTech multimodális adatbázis egyes, az ember–ember kommunikációra vonatkozó adatait. Az eredendően nyelvészek által kezdeményezett kutatás egyben láttatni engedi a nyelvtechnológia multimodális kiterjesztési lehetőségeit is.

Kulcsszavak: ember–gép kommunikáció, multimodalitás, korpuszépítés, annotáció, prozódia, szintaxis, pragmatika

1. Bevezetés

Évtizedekkel ezelőtt, amikor a számítógép először megjelent a környezetünkben, ez a környezet szakemberek (elsősorban matematikusok, egyéb természettudósok, csak később informatikusok) szűk csoportjára szorítkozott, azokéra, akik „értették a számítógép nyelvét". Eme exkluzivitás következtében a „kívülállók" a számítógépet valami megközelíthetetlennek, misztikusnak és olyannak tekintették, ami az ő világuktól merőben különböző, és szinte adottnak tekintették, hogy annak (meg)léte nem befolyásolja saját hétköznapjaikat. Pedig ha jobban belegondolunk, a számítógép is az ember teremtménye, így az embertől nehezen különíthető el, ráadásul olyan feladatokat bízunk rá, amelyek ilyen vagy olyan

mértékben, de mégis az ember szolgálatát jelentik. Egy hosszú evolúció eredményeként ezek a feladatok egyre inkább hétköznapi életünkhöz kezdtek közeledni, azaz egyre határozottabban érzékelhettük, hogy használatával „rólunk van szó" (egy levél megszerkesztésétől adóbevallásunkig vagy egy vonatjegy megvásárlásáig). Mivel most már nem csupán egy szűk kört érintő feladatról van szó (mint amilyen például egy termelési folyamat irányítása vagy az élve születések száma trendjének kiszámítása), mi, a hétköznapi felhasználók valóban úgy kívánunk tekinteni a számítógépre, mint ami emberi létezésünk része. Azaz – ha már bennünket szolgál – legyen a gép „gondolkodásában" is olyan, mint amilyenek mi vagyunk.

Nos, ez az a terület, ahol a technológiai fejlődés messze előtte jár a felhasználó hétköznapi élményeinek. Azt várjuk, hogy a gép szinte előre kiszámítsa vagy kövesse gondolatainkat, szándékainkat, ugyanúgy, ahogy ezt egy humán partnertől is elvárjuk. Természetesen azt is tudjuk, hogy egy humán partnerhez is alkalmazkodni kell, úgy, hogy lépéseinket a másik fél lépései (netalán megfejtett szándékai) is befolyásolják, azaz miközben elvárjuk, hogy a géppel való interakció számunkra könnyebb, emberközelibb legyen, ezen interakciónak mégis szigorú szabályokon kell alapulnia. Ezektől a szabályok azonban elvárhatjuk, hogy nekünk szóljanak, és így legfőképp az emberi kommunikáció szabályain alapuljanak.

De vajon milyenek az emberi kommunikáció szabályai? Vajon ismerjük őket? Úgy gondolhatjuk: persze hogy ismerjük, hiszen nélkülük nem tudnánk embertársainkkal sikeresen kommunikálni. De vajon meg tudjuk-e fogalmazni e szabályokat úgy, hogy a gép is megértse őket és így szinte természetes partnerünk lehessen? A kérdés nem is olyan egyszerű, miközben a megálmodott feladatok ezt kívánnák: azt, hogy például egy robot szobában elrejtett szemeivel-füleivel segítsen egy magányos arra rászorulót, vagy azt, hogy egy repülőtér forgatagában bennünket fáradhatatlanul és mindig ugyanolyan megbízhatósággal egészítsen ki egy keresett személy megtalálásában. Kövesse a hangunkat? A tekintetünket? A mozdulatainkat? Válassza ki az elérhető számtalan jel közül a relevánsakat, értse meg ezek összefüggéseit és ezekre alapozva jusson el egy döntés pillanatáig? Bizony, ezek az igények messze meghaladják azt a feladatot, amit egy ipari gép egyszerű vezérlő gombokkal történő irányítása jelent. Az ember hangsúlyozott részvételét az ember–gép interakcióban azon feltétel teljesülése mellett várhatjuk, ha mi, emberek megismerjük az ember–ember kommunikáció alapvető rendjét és ezeket a technológia számára is fogyasztható formában szabályokba foglaljuk, majd implementáljuk az ember–gép kommunikáció valamely technológiai alkalmazásában.

Az itt következőkben egy olyan projekt eredményeiről és jövőbeni terveiről számolunk be, amelynek a középpontjában az ember–gép kommunikáció technológiájának a továbbfejlesztése áll. E cél érdekében arra törekszünk, hogy jobban megismerjük az ember–ember kommunikáció alapvetőnek tekintett, ugyanakkor technológiai szempontból is releváns tulajdonságait, továbbá, hogy mindezen ismereteket olyan rendszerben reprezentáljuk, ami a technológia számára is elérhető lehet.[1] A projekt megtervezésekor kézenfekvőnek tűnt, hogy munkánk középpontjában a nyelv vizsgálata, ezen belül annak nyelvtechnológiai megközelítése álljon. Így céljaink között szerepelt, hogy a gyakorlatban akár közvetlenül is alkalmazható módon leírjuk és vizsgáljuk a spontán beszéd bizonyos akusztikai fonetikai tulajdonságait (F0, intenzitás, tempó), hogy a prozódia rendszeres leíásával hozzájáruljunk az automatikus beszédfelismerés és -szintézis tökéletesítéséhez. Ez utóbbihoz arra is szükség volt, hogy egy jelentős méretű korpuszon vizsgáljuk és gépi tanítás számára elérhetővé tegyük a spontán beszélt nyelvi szintaxis és a prozódia interfészét. Úgy gondoljuk, és első eredményeink is azt mutatják, hogy ez a spontán beszédre irányuló megközelítés a nyelvelméleten túl a nyelvtechnológia számára is számos új lehetőséget rejt magában. Bár projektünkben nem térünk ki olyan, jól megalapozott nyelvtechnológiai vizsgálatokra, mint az automatikus morfológiai és szintaktikai elemzés, korpuszunk multimodális jellege számos olyan izgalmas kérdés feltevésére ad lehetőséget, amelyek egyrészt kiterjeszthetik a nyelvtechnológia fogalmi és cselekvési körét, másrészt utat nyitnak további interdiszciplináris kutatásokhoz. Így a nem verbális gesztusok annotálása lehetővé teszi, hogy további, komplex adatokat szolgáltassunk a kommunikáció pszichofizológiai és kognitív vizsgálatához, különösen a spontán beszédben oly gyakori megszakadt vagy megszakított szintaktikai szegmensek és a gesztusok együtt járásának feltérképezésén keresztül. Ezt szolgálja a korpusz dialógusainak pragmatikai és funkcionális feldolgozása is, amelynek során a nyelvi, verbális elemeket szoros formális és funkcionális összefüggésben vizsgáljuk a nem verbális elemekkel. Az ilyen vizsgálatokra egy tágabb értelmű nyelvtechnológiának is nagy szüksége van, hiszen például egy avatárt nem elég „beszéltetni", hanem az adott kommunikatív helyzetben harmonizálni kell annak verbális és nem verbális viselkedését egyaránt.

Megközelítésünkben tehát a nyelvtechnológia tulajdonképpeni tárgya, a nyelv nem választható el a beszéd során megjelenő, azzal összefüggő egyéb, nem verbális jelektől, ami által a nyelvtechnológia is kiterjesztett értelmezést kap.

[1] A HuComTech munkacsoport munkája a TÁMOP 4.2.1 2008/9 projekt keretében indult és jelenleg a TÁMOP-4.2.1/B-09/1/KONV-2010-0007 projekt támogatásával zajlik. Egyes eredményei elérhetők a honlapján is: http://hucomtech.unideb.hu/hucomtech.

A jelen dolgozatban a HuComTech-team sokszálú tevékenységei közül a 2. pontban felvázoljuk azt az elméleti–technológiai modellt, amelynek kidolgozása mind eredménye, mind további alapja az ember–ember kommunikáció technológiai célú vizsgálatának. Ezt követően röviden bemutatjuk a létrehozott korpuszt, majd végigvezetjük az olvasót a különböző annotálási szinteken, bemutatva bizonyos kezdeti eredményeinket: a 3. pontban a vizuális szint, a 4. pontban az akusztikai szint (szöveg és prozódia) annotálását, az 5. pontban bemutatjuk egy, a nem szándékolt ismétlésekre irányuló vizsgálat multimodális adatokra támaszkodó eredményeit, a 6. pontban ismertetjük a spontán beszéd szintaxisának annotálását célul kitűző újszerű rendszert, majd a 7. és a 8. pontban, ugyancsak újdonságként, bemutatjuk egy tetszőleges kommunikációs esemény pragmatikai vonatkozásainak annotálását unimodális és multimodális megközelítésben.

2. A kommunikáció egy elméleti–technológiai modellje

A bevezető gondolatok szellemét követve egy technológiailag implementálható modellnek az ember–ember kommunikáció elméletén kell alapulnia. Olyan elméleten, amely kiindulásként figyelembe veszi a technológia, a gépi irányítás alapvető megkötöttségeit (ezáltal redukálva a fenti bevezető alapján szabadon ereszthető kívánságlistánkat) és ilyen megkötöttségek között keresi a humán kommunikáció szabályszerűségeit és írja le annak lehetséges működését.

A kommunikáció alapvető komplexitása megkívánja, hogy egy ilyen modell, miközben kellő általánossággal bír, figyelembe vegye az ember–gép interakció adott szándékolt alkalmazási területét. Egy ilyen kézenfekvő terület a humán–gépi interfészeké. Az ember–gép kommunikáció elméleti kutatásának a jelentősége is elsődlegesen a gépi interfészek kutatásában emelkedik ki (a teljesség igénye nélkül vö.: Wahlster 1991; Dix et al. 2003; Oviatt 2003). Ezek között Wahlster (1991) a kommunikáció komplex, dialógusalapú felhasználói modelljét írja le (XTRA rendszer), melyben – egyebek között – kimutatható a természetes és a gépi deiktikus jelek közötti különbség és igen tanulságos a gépi jelek használhatóságának kísérleti értékelése. Az interfészkutatáson belül külön hangsúlyt kap az interakció multimodalitása (vö. Flanagan 1997; Mariani 1997; Bernsen–Dybkjær 2005; Kuppevelt et al. 2005). A multimodális kommunikáció modelljeinek egy széles körű áttekintését adja Wahlster (2006), tanúsítva a modellek megfeleltetését az alkalmazások speciális céljainak. Az általunk is választott modellhez általános szemléletében Thórisson (2008) modellje áll közel, aki olyan absztrakt modulhierarchiákat tételez fel, amelyek kifejezetten a technológia (ezen belül a robotika) szempontjait veszik figyelembe.

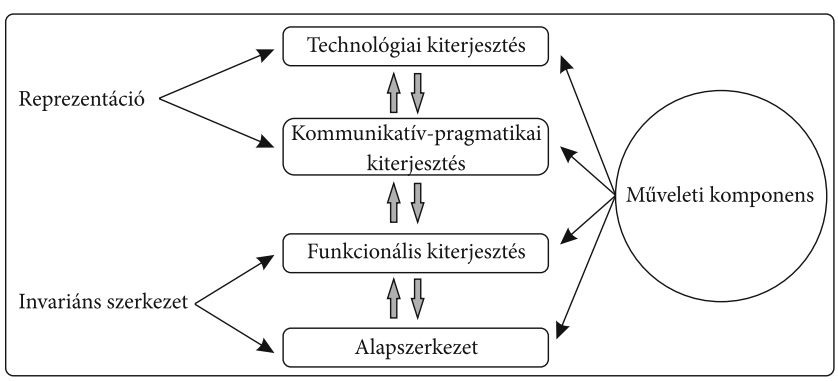
Az itt javasolt modell messzemenően figyelembe kívánja venni azokat a szempontokat (valójában kihívásokat), amelyeket a technológia állít egy, az ember–ember kommunikáció jellegét megközelíteni szándékozó alkalmazás elé. Így kiindulásként lényeges figyelembe vennünk, hogy a technológiai folyamatok lényegüket tekintve szigorúan szekvenciálisak. Ez azt jelenti, hogy egy esemény létrejöttének (a technológiai szabályozás egy bizonyos állapotának) mindig vannak előfeltételei és ezt követő következményei, úgy, hogy az ilyen szabályozás mindig egyirányú. Lényeges továbbá, hogy a szabályozás során mindig valamely egyetlen paraméter diszkrét értékének a beállítására kerül sor. Bár az lehetséges, hogy ennek során egynél több paraméter értékének a beállítása történjék meg egy időben, ez nem változtat azon, hogy a szabályozásnál egymástól jól elkülöníthető paraméterek értékei egyenkénti beállításáról van szó. Ilyen ételemben a szabályozás **moduláris.** Ezzel szemben a humán kommunikáció egyik fő jellegzetessége, hogy az egy időben bejövő jelek sokaságát (például köszönéskor az üdvözlő szavakat kéz- és testmozdulat, tekintet és fejtartás is kíséri, amelyek közül egyesek opcionálisak lehetnek) egyazon időben dolgozzuk fel, így a kommunikációt **holisztikusnak** érzékeljük. Ezért az elméleti–technológiai modellnek egyszerre kell egyrészt szekvenciálisnak és modulárisnak, másrészt holisztikusnak lennie. Ezen látszólagos ellentmondást azzal oldhatjuk fel, ha a modellünk szekvenciálisan moduláris, egyben multimodális lesz.

A humán kommunikáció főbb modelljei természetes módon nem a technológia követelményeire összpontosítanak, így eredményeik csak részben elérhetőek számunkra.[2] A **kódmodellek** (Shannon–Weaver 1949; Jakobson 1969) magára a kódhasználatra korlátozódnak, azaz arra, hogy a két információfeldolgozó eszköz közötti sikeres kommunikáció feltétele az azonos kód ismerete és a reprezentáció azonossága. E megközelítésben a kontextus nem játszik szerepet. A következtetéses modellek (Grice 1957; 1975; Lewis 1969) szerint a kommunikáció akkor sikeres, ha a kommunikációs partner meg tudja fejteni a beszélő által konkrét szituációkban, azaz kontextusokban létrehozott megnyilatkozások jelentését, azonban nem írják le kellően azt a folyamatot, amelynek során a kódhasználat beágyazódik magába a kontextusba. Az osztenzív-következtetéses modellek (Sperber–Wilson 1986/1995) már kezelik mind a kódhasználatot, mind a kódhasználat nélküli osztenziót és következtetést, azonban – hasonlóan az előző modellekhez – nem támaszkodnak a kommunikáció multimodalitására és szekvencialitására.

[2] Az itt említendő főbb modellek részletes vizsgálatára és az itt következő sorokban is felhasznált jellemzésére, továbbá a jelen tanulmányban javasolt elméleti-technológiai modellel való összevetésére l. Németh T. (2011a).

A javasolt modell alapját a generatív grammatika moduláris szemlélete képezi (vö. Chomsky 1965; 1977; 1986). Felteszi, hogy a kommunikáció meghatározó aspektusai megragadhatóak egymásra épülő modulokban zajló folyamatok leírásán keresztül. Ugyancsak felteszi, hogy az egyes modulok alapján egységesen ún. primitívek, tovább már nem bontható elemi részekből állnak és egy általános, a modulok fölött álló műveleti komponens ezen primitívekből összetett, immáron nem primitív szerkezeteket hoz létre. A műveleteket rekurzív módon alkalmazva elvileg végtelen összetettségű kommunikatív szerkezetek állhatnak elő. A legalsó szinten lévő modul, az invariáns Alapszerkezet a kommunikáció „szintaxisát" állítja elő, az erre épülő, ugyancsak invariáns Funkcionális kiterjesztés az Alapszerkezet által lehetségesnek (jól formáltnak, „grammatikusnak") tekintett formális szerkezetekhez ugyancsak lehetséges funkciókat rendel (azaz a „szintaxist" ellátja bizonyos mértékű „szemantikával"), végül a Kommunikatív-pragmatikai kiterjesztés képezi a reprezentáció szintjét, azt a szintet, ahol az immáron formailag és szűken vett értelemben funkcionálisan lehetséges (jól formált) absztrakt kommunikatív szerkezetek felszíni, azaz az adott kontextusban pragmatikailag aktualizált (az adott kommunikatív aktusra érvényesített) reprezentációt kapnak. Ezen, a modell elméleti komponense értelmében felszíni reprezentációnak valóságos, tárgyiasult felszíni reprezentációt a modell technológiai komponense, a Technológiai kiterjesztés ad, ahol a modell elméleti komponensének kimenete technológiai megfogalmazást nyer. Így tehát a Kommunikatív-pragmatikai kiterjesztés mintegy interfész-szerepet tölt be a modell elméleti és technológiai aspektusai között.

A modell sematikus felépítését mutatja az 1. ábra.



1. ábra. Az ember–gép kommunikáció generatív technológiai–elméleti modelljének sémája

A modell modularitásából és szekvencialitásából következik, hogy megvan annak az elvi lehetősége, hogy egyszerre szolgálja a szintézist (egy kommunikatív aktus létrehozását) és az analízist (egy észlelt konkrét aktus – funkcionális és pragmatikai – értelmezését), amit az ábrán a két irányú nyilak is sugallnak. Így a modell egységes elméleti–technológiai keretet biztosít arra, hogy a kommunikációban egy időben jelen lévő két, ellenkező irányú funkciót (a résztvevők főként váltakozó aktív és passzív szerepét – mint pl. a beszélő és a hallgató között) egyszerre ragadja meg és kezelje.

Néhány példa a modell egyes moduljainak alapjául szolgáló primitívekre:

Alapszerkezet: a kommunikáció kezdete, vége, fenntartása, időleges felfüggesztése és újrakezdése

Funkcionális kiterjesztés:

- strukturális funkcionális primitívek: a kezdés módja, a befejezés módja, a felfüggesztés módja, az újrakezdés módja,
- logikai funkcionális primitívek közé tartozik az állítás, tagadás, kérdés, kondicionális, kvantifikáció,
- holisztikus funkcionális primitívek: mellérendeltség, alárendeltség, fölérendeltség, a részvétel létrehozása és fenntartása, beszélőváltás, folyamatosság, érzelmek és szándékok

Kommunikatív-pragmatikai reprezentáció:

- *nem verbális*: vizuális, auditoros, vagy egyéb érzékeléssel, pl. érintéssel vagy szaglással közvetített formák, mint elemi mozdulatok, test- és arckifejezések
- *verbális*: szintaktikai (pl. állítás, kérdés, felkiáltás, óhajtás stb. kifejezése, valamilyen logikai funkció kifejezése, mint feltétel, kvantifikáció, negáció), lexikális (pl. szcenáriótól függő kifejezések), fonetikai–fonológiai (intonáció, beszédtempó, szünet)

Technológiai kiterjesztés:

- *interfész a modell elméleti oldala felé – markerek*: az Alapszerkezet, a Funkcionális kiterjesztés és a Kommunikatív-pragmatikai kiterjesztés primitívjeinek felszínen megjelenő, fizikailag mérhető megvalósulásai
- *interfész a modell technológai oldala felé – paraméterek*: a markerek technológiafüggő megvalósításának diszkrét bemeneti adatokat váró alapeszközei

A modulokra egységesen érvényes műveleti komponens olyan műveleteket tartalmaz, amelyekkel elvileg végtelen összetettségű kommunikatív szerkezetek állhatnak elő. Ilyen műveletek lehetnek: konkatenáció, iteráció, beágyazás, kiágyazás, közbeékelés, megszakítás, kombináció, és mindezek rekurzív alkalmazása.

A fentiekből kitűnik, hogy ahhoz, hogy egy kommunikatív eseményt akár szintetizálhassunk, akár analizálhassunk, minimálisan az szükséges, hogy az

adott esemény felszínen megjelenő összetevő szerkezeti elemeit, a markereket azonosítsuk. Ahhoz, hogy – a kommunikáció multimodalitásából kiindulva – ezen markerek együttállása alapján valamiféle kommunikatív értelmezést, valamint a technológia számára paraméterekben megfogalmazható vezérlést is adjunk, szükséges ezen markerek strukturális viszonyainak a meghatározása mind az eseménytípus, mind az aktuális esemény számára. Így az ember–ember kommunikáció tanulmányozása során az adatainkat egy olyan adatbázisba kell rendeznünk, amely alkalmas mind a kommunikáció általános jellegzetességeinek a meghatározására, mind az egyes konkrét esemény valamely konkrét mozzanatának az értékelésére. A HuComTech korpusz alapján létrejött adatbázis ezt a célt szolgálja. Az alábbiakban az ezen adatbázisból nyert adatok alapján rámutatunk az ember–ember kommunikáció során megfigyelhető bizonyos összefüggésekre. Ezen összefüggések feltárása vezethet el a bemutatott elméleti–technológiai modell technológiai alkalmazásához.

A HuComTech korpuszt egy 112 beszélő (egyetemi hallgatók) részvételével készült, összességében 50 órányi, beszélőnként fél-fél óra hosszú audió- és videófelvétel alkotja, amiből felvételenként kb. 5–5 perc felolvasás, 25–25 perc dialógus. A dialógusok során két személy spontán párbeszédét rögzítettük, egy formális és egy informális társalgási szcenárió keretei között. Az első (formális) dialógus egy szimulált állásinterjú, a második (informális) pedig egy, különböző témákat körbejáró irányított beszélgetés formájában valósult meg. A korpusz számítógépes feldolgozhatóságát a felvételekhez készült annotációk biztosítják, amelyek többféle megközelítésben (vizuális jelek, nyelvi egységek és kommunikációs események megfigyelése) címkézik fel a korpuszban vizsgált jelenségeket. Magában a korpuszban az egyes címkéket időjelekkel *(timestamp)* láttuk el, ami lehetővé teszi a kommunikáció számos jelenségének multimodális vizsgálatát és e jelenségek technológiai célú leírását. Az annotálást főként manuálisan végeztük úgy, hogy minden, előre megfogalmazott irányvonal mentén történt annotálás független ellenőrzésen ment át. Az adatbázis-lekérdezést felhasználtuk magának az annotálásnak az utólagos ellenőrzésére is. Az annotálás vonatkozott mind fizikai jellegű, de egyszerű leírást igénylő (pl. a tekintet iránya vagy kézi gesztusok jellege), mind interpretációt feltételező adatok (pl. érzelmi kifejezések) kinyerésére. Elkezdődött a korpusz automatikus annotálása is olyan területeken, ahol jól definiálható és igen pontos fizikai adatok kinyerésére van szükség (a beszédprozódia területén, elsősorban különböző felismerő algoritmusok számára).

3. A korpusz videoannotálásának elvei és egyes tapasztalatai

A fizikai jellegű, egyszerű leírást igénylő adatok (mint a tekintet iránya vagy kézi gesztusok jellege), valamint az interpretációt feltételező adatok (pl. érzelmi kifejezések) fentebb említett kinyerésére többfajta annotációt alkalmazunk.

A multimodális korpusz annotálása során kiindulásként kézenfekvő volt a videó- és az audiócsatornákat egymástól elválasztva kezelni. Ennek elsődleges oka az, hogy egy technológiai alkalmazásnál a kétféle csatornából érkező jeleket természetük különbözősége folytán különböző eszközökkel (szenzorokkal) tudjuk érzékelni, de az is, hogy egy szintézis során ugyancsak külön-külön kell létrehozni videó- és audiómintázatokat. A videó kézi annotálásánál egyrészt, mint mondtuk, diszkrét, fizikai természetű adatokat figyeltünk meg és annotáltunk, másrészt egyes mintázatokhoz interpretatív jegyeket rendeltünk. Az annotálás eme kettőssége (deskripció és interpretáció) előrevetítette az így kinyert értékek kettős felhasználását: ezen adatok megléte lehetőséget teremt a szintézis (fizikai elemek egyenkénti manipulálásával történő) technológiai megvalósítására és – ugyanezen jegyek együttállásának vizsgálatával – a kommunikatív esemény multimodális értelmezésére. (A videoannotáció ezen utóbbi mozzanata összetettebb szinten visszaköszön a multimodális pragmatikai annotáció esetében; vö. 8. pont.)

A projekt keretében fejlesztett Qannot annotáló program (Pápay et al. 2012) lehetővé teszi, hogy egy videót a lehető legkisebb részekre *(frame)* bontva megfigyeljük az esemény lelfolyásának videómozzanatait és az észlelt jelenségek időtartamát (kezdetét és végét), valamint annak milyenségét megfelelő címkével jelöljük. A különböző szempontoknak megfelelően az annotálás különböző szinteken történt.

A videoannotáció szintjei három csoportba sorolva (az angol nevek értelmezése utánuk zárójelben áll) az alábbiakban láthatók:

1) basic (a videófelvételt technikailag azonosító szintcsoport):

- comevent class (itt jelölendő a teljes kommunikációs eseménysor, az egész interjú kezdete és vége)

2) physical (a videoannotációban fizikai természetű jegyek alapján azonosítható szintcsoport):

- facial expression class (az arckifejezés szintje)
- gaze class (a tekintet iránya)
- eyebrows class (a szemöldök mozgása)

- headshift class (fejmozgás)
- handshape class (a kézfejek formája)
- touchmotion class (adott testrész érintése)
- posture class (testtartás)
- deictic class (utalás valamire kézzel)

3) functional (a videoannotációban funkcionális természetű jegyek alapján azonosítható szintcsoport; ezeken hallható az alany hangja is):

- emotion class (érzelemkifejezés)
- emblem class (ez az ún. emblémaszint az egyetértés és az egyet nem értés szintje)

Anélkül, hogy az annotálás folyamatát részletesebben ismertetnénk (erre l. Földesi 2011, 39), a következőkben néhány kiragadott példával érzékeltetni kívánjuk a videofelvételeken észlelhető modalitáskombinációk sokrétűségét, úgy, hogy párokat vagy hármasokat adunk meg, amelyeknek a tagjait egy arckifejezés-típus és egy tekintetirány, vagy az előbbi kettő és esetleg egy jellegzetes kéztartás képezik. Először az interjúkban elhangzott üzenetek információtartalmához való hozzáállás többsíkú (több modalitásban történő) kifejeződését vizsgáljuk, de tovább is óhajtunk lépni annyiban, hogy rátérünk a még nem kezelt, egy unimodális annotáció keretein belül is kérdéses mozdulatokra vagy mozdulatsorokra, címkével nem jelölhető arckifejezésekre. Tehát olyan, az adott megfigyelt személyre (interjúalanyra) jellemző, az egész formális vagy informális interjú alatt visszatérő mozdulatsorokról, arckifejezésekről stb. is lesz szó, amelyekhez nem rendelhető funkció.

3.1. A szegmentálás és következményei

A videoannotáló program segítségével a korpuszunkban rögzített interjúanyagot a benne fellelt kommunikációs események részmozzanataira daraboljuk fel. A feladat lényege a videofelvételek alatti, annotációs szinteket tartalmazó szalagon (egyenlő egységekre – *frame*-ekre – tagolt elemző felületen), időhatárok kijelölése után, minden határozottan észlelhető történés megcímkézése.

Az annotátorok munkájuk végeztével elérik, hogy ne csak maguk a videófelvételeken látott események és részmozzanataik legyenek láthatóak, illetve láthatóbbak, mint annotáció nélkül, hanem ezek időbeli kiterjedése is, ami a kész annotációkról leolvasható.

Vigyáznunk kell azonban, hogy mekkora szeletekre aprítunk fel egy mozdulatsort annak felcímkézéséhez. Mozdulatsoron az egymást követő azonos, hasonló vagy különböző mozgásokat értjük. Nincs okunk rá, hogy a mozdulatsor egyes részeit feltétlenül egymáshoz tartozónak vegyük.

Fontos megkülönböztetni egyszeri és periodikus mozgást is. Ugyanis előfordul, hogy a megfigyelt személy nem határozott, teljes mozdulatot tesz, hanem csak megkezd egy mozdulatot, azaz mozgást végez, de mozdulatot nem hajt végre.

Egyszeri (határozott) mozdulat a meghatározott számú *frame*-en keresztül (hosszabb időn át) végzett, az adott testrészt a nyugvópontjára visszatérítő mozgás, pl.: amikor a személy a címzettre (kérdezőre) mutat, vagy osztenzív mutatás történik, mert a beszélő a felemelt kezét visszaviszi a törzse mellé vagy ismét a combján nyugtatja, esetleg ujjait megint összekulcsolja.

Egyszeri mozgás a pillanatnyi időre megemelt kéz vagy láb, a testtagok rándulásai: ezeket ugyan annotálhatjuk, de felesleges funkciót tulajdonítani nekik.

Periodikus mozgás például a lábrázás vagy -lóbálás, ahol a mozgás, amely ismétlődik, sokszori és oly rövid, hogy megcímkézése legfeljebb felületesre sikerülhet, mivel nem feleltethető meg neki az „egy *frame* jobbra, egy *frame* balra" – különben rendkívül körülményes – címkézési módszere.

Abban az esetben, ha az illető csak az ujjával malmozik, ez a módszer még célravezető lehet, és a mozdulatsornak szó szerint teljes (címkékkel való) lefedettséget biztosít. A valódi, azaz gyors periodikus mozgás ellenben saját címkét venne igénybe, mivel egy frame-en belül a mozgás többszöri megismétlése is lehetséges (egy frame a Qannotban kb. 247 millisecundum).

A túl aprólékos vagy éppen felszínes szegmentálás másik következménye az, hogy igen rövid, egy vagy két egységnyi időszakasz alatti történések nem ugyanazokat a címkéket kapnák, mint amelyeket egy fokozott sebességű annotáció alatt. Megfordítva: amennyiben az annotátor mindig hosszú, több perces szakaszokat határol el, a kommunikációs események és az őket alkotó mozdulatok, mozdulatsorok vagy a változó arcjáték részei nem különülnek el megfelelő élességgel. Feltétlenül igaz viszont az, hogy nem szabad (a fentiek ismeretében nem is lehet) részeire tagolni egy periodikus mozgást, mint amilyen a fej ingatása (címkéje: „sideways"), a bólogatás („nod") vagy a fej rázása („shake"). Nyitva hagyható az a kérdés, hogy a fej csóválása (a rázásnál jóval lassabb jobbról balra és balról jobbra fordítása) periodikus mozgásnak számítson-e. Ha igen, ugyanúgy „shake" címkével kell ellátni; ha nem, jobbra fordításnak („turn right") és balra fordításnak („turn left") kell annotálni.

A periodikus mozgások gyakorta a pótcselekvéseknek feleltethetők meg, bár a kéz tördelése („broke") kivétel.

3.2. Az arcon és a hanggal közvetített érzelmek lehetséges funkciói

A videoannotálás során arckifejezéseket is annotáltunk. Az érzelmi állapotokra utaló címkékkel az arcon közvetített helyzetértékelést is rögzítjük: egy szorongó vagy gunyoros arckifejezésből nagyjából meghatározhatjuk a beszélő viszonyulását saját mondanivalójához vagy a partnerétől hallott kérdéshez/kommentárhoz. Azonban az annotátor közelítőleg sem állapíthat meg semmit a beszélői attitűdről, ha nincs tudatában, hogy a mimika – főleg a formális beszélgetésekben – függetlenedhet a megfigyelt személy tényleges hangulatától és beállítódásától, éppen valódi véleményének eltitkolása okán. Ezért az érzelmi vonatkozású videoannotációs szinteken („facial expression class", „emotion") adott időszakaszon mindig a korábbi állapothoz vagy a későbbihez viszonyítva adhatunk címkét (azaz az éppen annotálandó szakasznál nagyobb szakasz figyelembe vételével), ugyanis valaki akkor „happy", ha a későbbi vagy korábbi nyájas mosolyához képest boldog. Bár ez az eljárás erősen szubjektívnek tűnhet, de vegyük figyelembe, hogy mindenkinek van egy „alapmimikája", vagyis a helyzethez illesztett állandónak nevezhető arckifejezése, amelyhez mint alapállapothoz a felfokozottabb érzelmi állapotokból visszatér. Ez pedig ugyanúgy lehet egy közönyt sugárzó merev arckifejezés, mint egy folytonos udvarias mosoly. Az alapmimika és a tényleges érzelmi állapotot tükrözni akaró közötti szembenállás tagadhatatlanná lesz, ha összevetjük a „facial expression class" és „emotion" annotációs szinteket. Az „emotion" szinten, az egyén hangját hallva meggyőződhetünk róla, arckifejezése mennyire őszinte, illetve, amit nyugodt vagy derűs hangja nem árul el, az esetenként az arcára van írva. Példák:

(1) Fájlazonosító: `071_J_C2.mts`

Vizsgált időszakasz (perc, másodperc, századmásodperc): 03:13:20 – 03:14:75

Elhangzott szövegrész: „minden további nélkül megtanulom."

Megjegyzés: A megfigyelt személy arca („facial expression class") előbb nyugodt („natural"), aztán szomorú („sad"). Ugyanezen időben az érzelemkifejezés szintjén („emotion") feszült („tense") van jelölve.

(2) Fájlazonosító: `071_J_C2.mts`

Vizsgált időszakasz: 03:25:59 – 03:32:75

Elhangzott szövegrész: „…keretein belül is bizonyos technológiákkal."

Megjegyzés: A vizsgált időszakaszban az egyén eleve feszült, de az eztán következő idegességéhez („tense") képest ez még semleges érzelemként („natural") jelölhető az érzelemkifejezés („emotion") szintjén.

A 071-ből vett második példa nem tekintendő együttjárásnak, a viszonyító annotációra viszont jó péla lehet.

3.3. Együttjárások

Az interjúban elhangzottakat a videókon megerősítheti vagy relativizálhatja a taglejtéssel kifejezett kommunikatív funkció, amely az annotáció két szintjén jelölhető. Jelölhető mint utalás („deictic" level), és/vagy jelölhető az ún. emblematikus szinten („emblem"). A partner figyelmének felkeltése például megjelenhet valamilyen gesztussal, a partnerre (a kérdezőre) figyelés megjelenhet gesztus nélkül. Ebben az annotációtípusban csak egy címke van a figyelemre („attention"). Az unimodális (csak lehetséges kommunikatív funkciók jelölőit nyilvántartó) annotációban elfogadott, videoannotációban viszont paradoxnak hat, hogy egyetértő hangzó megnyilatkozást fejrázás kísérjen (a példákban az időtartam jelölése: perc:másodperc:ezredmásodperc).

(3) Fájlazonosító: `071_J_C2.mts`

Vizsgált időszakasz: 03:13:20 – 03:14:75

Elhangzott szövegrész:„minden további nélkül megtanulom."

Megjegyzés: A felsorolásban emblémának nevezett szinten az elhangzott szövegrész értelmében egyetértés („emblem: agree") jelölhető, ami viszont fejrázással jár együtt („headshift: shake").

(4) Fájlazonosító: `077_J_C2.mts`

Vizsgált időszakasz: 01:26:79 – 01:28:76

Elhangzott szövegrész: „hát szeretnék."

Megjegyzés: Az elgondolkodó/emlékező arckifejezés („facial expression class: recalling") a másfelé (nem előre) nézéssel („gaze: left down") jár együtt.

(5) Fájlazonosító: `077_J_C2.mts`

Vizsgált időszakasz: 01:22:79 – 01:26:76

Elhangzott szövegrész: „miért jelentkezett a hirdetésre?" (A kérdező mondja, a kérdezett ezalatt figyel.)

Megjegyzés: A figyelem („attention") az annotátor számára nem pusztán az előre irányuló merev tekintetből tevődik össze („emotion: natural"; „gaze: forwards"). Ha így lenne, nem kellene a figyelmet, amit észlel, az arra fenntartott szinten jelölni („emblem: attention").

(6) Fájlazonosító: `077_J_C2.mts`

Vizsgált időszakasz: 01:42:00 – 01:44:35

Elhangzott szövegrész: „… vagy távolállónak érez?" (A kérdező mondja, a kérdezett figyel.)

Megjegyzés: A 077_J_C2.mts videón (ahogy sok másikon is) megfigyelhető, hogy amíg a figyelem a kérdező felé fordul, az érzelemkifejezés az arcon felfüggesztődik. Ez az erős koncentráció jele.

(7) Fájlazonosító: `077_J_C2.mts`

Vizsgált időszakasz: 00:40:39 – 00:42:75

Elhangzott szövegrész: „uhum. és ilyen fodrászképző?" (A kérdező mondja, a kérdezett figyel.)

Megjegyzés: A vizsgált időszakaszban figyelem és feszült arckifejezés észlelhető.

(8) Fájlazonosító: `077_J_C2.mts`

Vizsgált időszakasz: 01:29:20 – 01:29:96

Elhangzott szövegrész: „… állást kapni."

Megjegyzés: Figyelem („attention") az emblémaszinten és meglepettség („surprise") az arcon.

(9) Fájlazonosító: `077_J_C2.mts`

Vizsgált időszakasz: 01:38:40 – 01:39:15

Elhangzott szövegrész: „… kérdés, …"

Megjegyzés: Ha a figyelem lankad, a kérdezett lefelé néz (jobbra le, balra le vagy középre le). Ebben a konkrét esetben a kérdezett balra lefelé néz („gaze: left-down"). 01:38:40-ig figyelem („attention") volt jelölhető, ettől az időponttól kezdve már nem, vagyis a figyelem megszűnt.

(10) Fájlazonosító: `077_J_C2.mts`

Vizsgált időszakasz: 01:45:60 – 01:47:56

Elhangzott szövegrész: A kérdezett semmit sem mond, de a fejét rázza („headshift: shake").

Megjegyzés: Itt szemlesütés figyelhető meg („gaze: left down"), valószínűleg kínos meglepetés hatására („facial expression: surprise").

(11) Fájlazonosító: `077_J_C2.mts`

Vizsgált időszakasz: 01:44:79 – 01:45:56

Elhangzott szövegrész: „hát"

Megjegyzés: Ugyanaz a szemlesütés látható, mint az előző példában, viszont az arckifejezés gondolkodó, emlékező, egy szóval elmélkedő („recalling").

Azt is megjegyezzük, hogy a példákban nem biztos, hogy a megadott arckifejezés nem tart tovább az időpontokkal leírt szakasznál, de bizonyos, hogy az együttjárás az így megjelölt időintervallum teljes hossza alatt fennállt.

A képi anyag, illetve hangos képanyag annotációja után most a csak a hanganyag annotálásával szerzett tapasztalatainkról számolunk be.

4. Az audioannotálás tapasztalatai. Automatikus prozódiai annotáció

Az audióanyag első annotálását kézzel végeztük. Ennek folyamán létrejött az elhangzott szöveg átírása, továbbá sor került bizonyos kommunikatív jelenségek (egyebek között beszélőváltás, megakadás, újraindítás, különböző érzelmek) annotálására. E kezdeti szakaszban mindezeket a kommunikatív jelenségeket olyan, időnként esetlegesnek tűnő időintervallumon belül határoztuk meg, amely sokszor, de távolról sem mindig a tagmondat valamely tentatív fogalmának felelt meg. Ezen intervallumon belül a jelenségek pontosabb időbeli meghatározására nem került sor. Míg ez az annotáció számos fontos kommunikatív jelenség jelölését lehetővé tette, az az igény, hogy mindezek strukturálisan és időben is pontosabban köthetőek legyenek egy formális nyelvi szerkezethez (és így adatbázisban az együttállásokat is figyelembe véve lekérdezhetőek legyenek), a későbbiekben szükségessé tett egyrészt egy formálisan megragadható szintaktikai annotációt, másrészt azt, hogy a prozódiai paramétereket egy speciális fonetikai szoftver alkalmazásával automatikusan annotáljuk. Az alábbiakban ez utóbbi annotáció részleteit ismertetjük.

4.1. Az automatikus prozódiai annotáció szerepe

A multimodális pragmatikai annotáció során feltárt kommunikációs események, illetve a szintaktikai annotációkban jelölt nyelvi egységek elemzési lehetőségei csak úgy válhatnak teljessé, ha gépileg feldolgozható formában rendelkezésre állnak az őket kísérő, azok potenciális markereit képező vizuális és akusztikus jegyek. A HuComTech korpusz videoannotációi a vizuális oldalról szolgáltatják ezeket az információkat (nem-verbális gesztusok, tekintetirány, különböző arckifejezések manuális és automatikus címkézése). Az automatikus prozódiai annotáció célja, hogy hasonló információk akusztikai oldalról is elérhetővé váljanak.[3]

A megvalósítás első lépését a beszédfolyam alapvető fizikai paramétereit képező adatok (mint amilyen az alapfrekvencia és az intenzitás) kinyerése és tá-

[3] Az automatikusság fokozott igénye ezen a területen nem csupán a hatékony megvalósítás érdekében áll fenn, hanem egyúttal bizonyos szupraszegmentális jelenségek (mint pl. a relatív beszédtempó vagy hangmagasság) manuális leírási nehézségeit, megbízhatatlanságát és szubjektivitását igyekszik áthidalni. A nyelvhasználat során feltehetőleg öntudatlanul, de operálunk ezekkel az információkkal, viszont a prozódiai annotáció esetében (ahol az annotátor szubjektív észleleteire reflektál) ugyanezen jelenségek sokkal kevésbé ragadhatók meg egyértelműen és objektív módon, mint például a fejmozgás iránya a videó annotációja során.

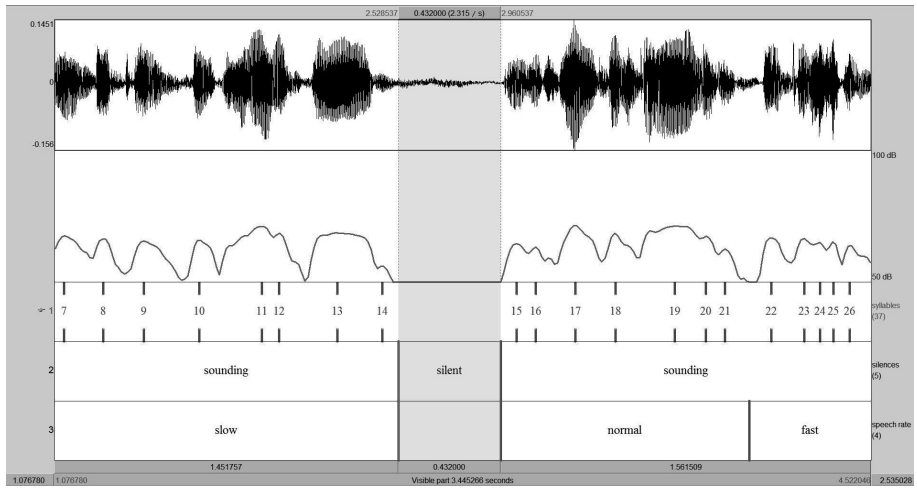
rolása jelenti, amely után közvetlenül is lehetőség nyílik a különböző pragmatikai és szintaktikai címkék mentén történő vizsgálatok elvégzésére. A szükséges adatok lekérdezésével statisztikai számításokat végezhetünk például arról, hogy egy adott típusú kommunikációs gesztus a többihez képest milyen átlagos hangmagassággal és intenzitással valósul meg a verbális közlés folyamán. A következő lépést az intenzitás- és frekvenciaadatok összetettebb feldolgozása jelenti, amelynek során további szupraszegmentális jelenségeket kívánunk lekérdezhető formában felcímkézni. Ezeknek a címkézési eljárásoknak képezi tárgyát a beszédtempó és a beszéddallam annotációja, amelynek lépéseit az alábbiakban egy kapcsolódó tanulmány (Szekrényes et al. 2011) alapján összegezzük.

4.2. A beszédtempó annotációja

A beszédtempó annotációjához elsősorban a beszédfolyam egy olyan automatikusan detektálható elemére van szükségünk, amelynek egy adott időegységre mért gyakorisága, sűrűsége megragadhatóvá teszi annak metrikus (időbeli) struktúráját. Jong és Wempe (2009) a beszédtempó vizsgálatához a szótagmagokat választották mérési objektumként, amelyeknek az automatikus detektálása a beszéd intenzitásának dinamikus kiugrásai alapján, az intenzitásgörbe csúcsértékeinek meghatározott küszöbértékek (csúcsok közötti minimális értékbeli különbség stb.) szerinti szűrése által történik. A beszéd sebességének ingadozása így az intenzitáscsúcsok közötti távolság változásain keresztül válik mérhetővé, ahol minden intenzitáscsúcs egy szótagmag helyét reprezentálja (2. ábra). A szótagmagok detektálása után előzetes kalkulációkat végezhetünk az adott beszélő egyedi beszédtempójáról, majd a beszélő normál (átlagos) beszédtempójához viszonyítva osztályozzuk az aktuális beszédtempót a felcímkézendő beszédszegmensek mentén. A címkézési eljárás egy lehetséges kimenetét a 2. ábra szemlélteti a Praat program (Boersma – Weenink 2005) annotációs felületén. A módszer tervezett implementálásának további részleteiről lásd még Szekrényes et al. (2011) tanulmányát.

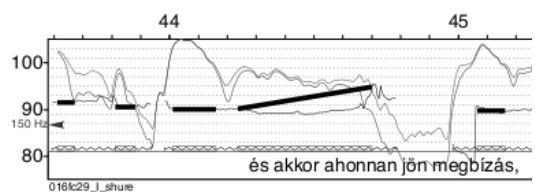
4.3. A beszéddallam annotációja

A prozódiai annotáció következő lépését az alapfrekvencia progressziójának (a dallammenet alakulásának) elemzése és címkézése jelenti. Az eljárás során a beszédfolyam meghatározott szegmentumaiban kimért F0 értékek leírta dallammenetet igyekszünk a progresszió fő irányvonalait megragadó stilizált formára



2. ábra. A beszédtempó annotációja

hozni (ezt szemléltetik a 3. ábra vastag vonalai), amelyhez meghatározott küszöbértékek használatával valamilyen egzakt tonális karaktert jelelő annotációs címke (emelkedő, ereszkedő, eső stb.) vagy címkekombináció rendelhető.



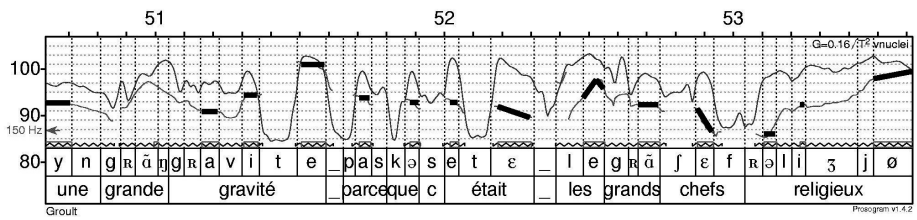
3. ábra. A beszéddallam stilizációja

Ennek a célnak a megvalósításához gyakorlati és elméleti útmutatásként Piet Mertens munkáját (Mertens 2004) terveztük felhasználni, aki tanulmányában számos fontos előzetes kikötést fogalmaz meg a prozódiai annotációval kapcsolatban:

1. Az annotációnak alapvetően az ember által érzékelhető intonációt kell reprezentálnia objektív és könnyen értelmezhető módon.
2. Az alapfrekvencia változását hosszabb beszédfolyamon keresztül is tükröznie kell a szélesebb tartományokra kiterjedő változások rögzítése érdekében.
3. A fizikai jelek időbeli szerveződését meg kell őriznie a szünetek, hezitációk, beszédtempó és a ritmus azonosíthatósága érdekében.
4. Az annotációnak automatikusnak vagy félautomatikusnak kell lennie.

5. Az annotáció elméletsemleges kell, hogy legyen, a széleskörű használhatóság érdekében.
6. Az annotáció lehetőleg időben illesztett fonetikai és szöveges átírást tartalmazzon az olvashatóság és szöveges keresés lehetőségének biztosítása érdekében.

A szerzők által kifejlesztett transzkripciós rendszer (l. 4. ábra) a vokális szótagmag alapfrekvenciájának stilizált kontúrját felhasználva automatikusan vagy félautomatikusan rendel prozódiai annotációt fonetikai transzkripcióhoz. Maga a stilizálás a fönti kikötésekkel ellentétben ugyan nem teljesen elméletsemleges, hiszen a tonális érzékelés pszichoakusztikai modelljére épül (l. Alessandro–Mertens 1995), viszont megőrzi az akusztikai jel temporális jellemzőit, és beépíti a szöveges, illetve a fonetikai átírást is, ahol ez utóbbi a vokális szótagmag azonosításában játszik szerepet.



4. ábra. A Mertens-féle transkripciós rendszer

A módszer algoritmizálása a Praat beszédfeldolgozó program beépített szkriptnyelvén történt. A transzkripciókat grafikus formában generáló Praat szkript a hozzá tartozó dokumentációval együtt Prosogram (v2.8) néven szabadon hozzáférhető,[4] így a jelen tanulmányban részletesen nem ismertetjük. Az eredményül kapott stilizációkat a különböző beszédszegmensek dallamkarakterének címkézéséhez kívánjuk felhasználni. Mivel a program grafikus formátumú kimenete az ehhez szükséges információk feldolgozását nem támogatja, az algoritmus olyan technikai jellegű átdolgozása szükséges, amelynek eredményeként a kimenet numerikus formában tartalmazza a stilizációk kezdő- és végpontjának időpillanatait és frekvenciaértékeit is.

A munka jelenleg előkészítő szakaszban van. Az automatikus prozódiai annotáció sikeres implementációja után lehetőség nyílik olyan lekérdezések összeállítására, amelyek a manuálisan annotált kommunikációs események prozódiai markereinek feltérképezését segíthetik elő. Ezeknek a feltárt prozódiai jellemzőknek a birtokában a későbbiekben lehetővé válik a vizsgált kommunikációs események (társalgási fordulók, témaváltások stb.) gépi detektálásának vagy pre-

[4] http://bach.arts.kuleuven.be/pmertens/prosogram/

dikciójának algoritmizálása, amely a számítógéptől ugyanezen prozódiai jellemzők felismerését követeli meg. A prozódiai annotáció során kifejlesztett eljárások ehhez szintén jól alkalmazhatóak lesznek.

5. Esettanulmány: a nem szándékolt ismétlések multimodális jegyei

Az előző pontokban ismertetett video- és prozódiai annotálás után egy rövid esettanulmányban mutatunk ízelítőt a HuComTech korpuszban rejlő lehetőségekből. A korpusz adataira támaszkodva azt a feltevést kívánjuk alátámasztani, hogy a nem szándékos ismétlések egy, az ismétlést követő tartalmas szó[5] memóriából történő előhívását könnyítik meg. Ezt a szót a beszélő nyomatékosítani kívánja, hogy ezzel is segítse a feldolgozást a hallgató számára. A beszélő a nyomatékosításhoz nemcsak akusztikai jelenségeket, hanem vizuális, nem verbális markereket is alkalmaz. Korpuszunk lehetővé teszi, hogy ezeket a nem verbális jeleket együtt vizsgáljuk az akusztikai jelekkel, így a nem szándékos ismétlés multimodális vizsgálata hozzájárulhat a kommunikáció igen összetett jelenségének jobb megértéséhez és az ismeretek számos nyelvészeti, nyelvtechnológiai és egyéb célú alkalmazásához.

Az ismétléseket úgy tekinthetjük, mint a spontán beszéd folyamatában, egy szó kivitelezésének a beszélő bizonytalanságából adódó, nem szándékos megismétlését (Gósy 2002). Az ismétlés történhet változtatással vagy változtatás nélkül (Fox et al. 1996, 230). Jelen esettanulmányban az egyszerű, változtatás nélküli ismétléseket tanulmányozzuk.

5.1. Az esettanulmány vizsgálatának előzményei

Az ismétléseket szintaktikai környezetükben, az azokat tartalmazó tagmondatokban vizsgáltuk. Az alábbiakban, az egyszerűség kedvéért, tagmondatnak nevezzük azt az egységet, amelyben az ismételt szó és az ezt követő tartalmas szó elhelyezkednek, függetlenül attól, hogy az adott tagmondat a spontán beszédbeli megnyilatkozásban gyakran jelentősen különbözhet attól, amit a leíró nyelvtan

[5] Kenesei (2000) tartalmas szavaknak nevezi a főnevek, igék, melléknevek, határozószók alkotta nyitott szóosztályokat, amelynek komponensei kölcsönzéssel, szóképzéssel stb. gyarapíthatók, míg ezzel szemben a funkciószavak (segédigék, kötőszavak, névelők) száma változatlan, állandó (*op.cit.* : 95).

ezen fogalma takarna.[6] A tagmondatokat a diskurzusbeli elhelyezkedésük szerint csoportosítottuk: forduló fenntartása, fordulóváltás (átadás/átvétel).

A HuComTech korpusz adatai alapján történt korábbi vizsgálatok (Abuczki 2011; Tóth 2011), melyek célja a fordulóátadás, illetve -átvétel során megjelenő tekintet- és gesztusmintázatok elemzése volt, azt találták, hogy fordulóátadáskor az interjúalany tekintetét beszédpartnerére szegezi, és a forduló átvételekor is hasonló tendencia figyelhető meg. Ahhoz, hogy a vizuális jegyek értelmezésekor kizárjuk a fordulóátadás és -átvétel ilyen hatását, a nem szándékos ismétléseknek csak azon eseteit vizsgáltuk, amelyek nem estek egybe fordulóváltással. Az ilyen esetek közül kizártuk azokat a tagmondatokat, amelyekben együttbeszélés fordult elő, ugyanis ennek pszicholingvisztikai mechanizmusa eltérhet az egyedülbeszélésétől (vö. Gósy 2008; Gyarmathy 2011).

A HuComTech annotált spontánbeszéd-korpuszból felhasznált 6 informális dialógus korpuszba rendezett, időben szinkronizált multimodális adatai lehetővé teszik, hogy segítségükkel az ember–ember interakció vizuális és auditív markereit együttesen elemezhessük, együttállásokat és összefüggéseket állapíthassunk meg verbális és nem verbális jegyek között. Ezáltal hozzájárulhatunk a nem szándékolt ismétlések multimodális tulajdonságainak a megismeréséhez, és ezek felismeréséhez akár egy ember–ember, akár egy ember–gép kommunikációra fókuszáló alkalmazás számára. Az itt ismertetendő eredményeink a vizsgált dialógusokon belül az interjúalanyok adataira támaszkodnak.

Az interjúalanyok 97 tagmondatában fellelhető 106 ismétlést, valamint az azokat közvetlenül követő tartalmas szavaknak a beszélő általi szándékolt előhívására utaló tekintetét és gesztusait vizsgáltuk abból a szempontból, hogy e kétféle mozzanat során az alanyok milyen vizuális markereket használtak.

5.2. Az egyes markerek együttállása: az interjúalanyok tekintetviselkedése az ismétlés során

A beszélők a szó ismételt kiejtése során 26,4%-ban (28 esetben) tekintettek a szemben ülő partnerre (továbbiakban röviden TP (= tekintet a partnerre)), és 73,6%-ban (78 esetben) egy ettől eltérő irányba (rövidítve TM (= tekintet más irányba)) irányították tekintetüket.[7] Ez az ismétlés alatti magas arányú tekintet-irány-változás feltehetően azért történt, mert eközben a megformálandó szón

[6] A spontán beszéd szintaktikai felosztásáról részletesebben a következő pontban lesz szó.

[7] Bár lekérdezhetőek korpuszunkból, az irányok részletezésére (bal-jobb, fel-le) itt nem törekedtünk, mert személytől függ, ki milyen irányba tekint a beszéd során.

vagy szintaktikai szerkezeten gondolkodtak, azaz a tekintet irányának módosítása összefüggésben volt a nem szándékolt ismétléssel. Kitűnt, hogy azokban az esetekben, ahol az interjúalany tekintete az ismétlés során az interjúvezető felé fordult, a beszéd menetében nem történt megtorpanás. Ebből arra következtetünk, hogy ekkor a soron következő szó/kifejezés előhívása és produkciója minden bizonnyal nem okozott különösebb kognitív nehézséget.

5.3. Az ismétlések utáni tartalmas szavak tekintetmintázatai

Az esettanulmányhoz felhasznált adatokat két csoportba sorolhatjuk annak alapján, hogy az ismételt szót követően megtörténik-e a tartalmas szó előhívása:

1. Megtörténik az előhívás: a beszélő ismétlését követő első tartalmas szó alatti tekintetmintázatok vizsgálhatók. Például:[8]

(12) hogy mennyire változatos a <a> <a> <a> kultúra

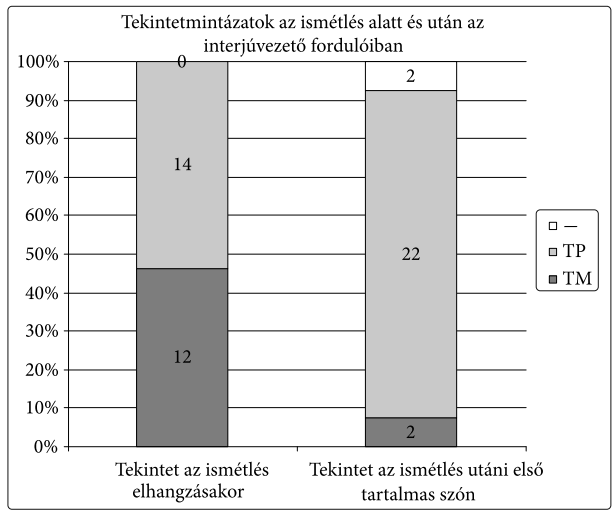
2. Nem történik meg az előhívás (∅ = nincsen ismétlést követő tartalmas szó, amelynek a tekintetmintázatát vizsgálhatnánk). Vagy azért, mert az ismétlést nem követi tartalmas szó, mert az interjúalany új szintaktikai egységet kezd, ezáltal befejezetlenül hagyva a megkezdett tagmondatot (13), a tagmondatok közötti határt a „|" jel jelzi), vagy mert tartalmas szó helyett újabb ismétlés következik (14). Ezekben az esetekben – a tartalmas szó előhívásának hiánya miatt – ezen mozzanat nem vizsgálható:

(13) szóval ez így <így> – | ah, ez így fáj.

(14) úgyhogy <úgyhogy> én <én> amikor így <így> láttam,

Az esetek 90%-ában (96 db tartalmas szónál) vizsgálható volt a tekintet iránya az ismétlést követően, így ezeket az első csoportba soroltuk. Az első csoportba tartozó mintákat is megvizsgáltuk a tekintet iránya szerint (TM/TP), és a következő eredményre jutottunk: a tartalmas szavak 25%-ában nem tekintettek a partnerre a beszélők (TM = 24 db), míg a vizsgált esetek 75%-ában a tekintet a partner felé irányult (TP = 72 db). Az 5. ábra tartalmazza az ismétlések kétféle tekintetmintázatát, valamint az első tartalmas szavak tekintetmintázatát, illetve annak hiányát darabszámok szerint. A második csoporthoz sorolható minták az esetek 10%-át (10 db) alkotják. Mivel ez utóbbi esetben a tartalmas szó elmarad, a lehetséges tekintetmintázatra és annak vizsgálatára sem került sor.

[8] Az itt használt jelek megegyeznek a HuComTech korpusz audioannotációja során használt szimbólumokkal, azaz: „%" = nyújtás (a megnyújtott hang előtt szerepel), „< >" = ismételt szó, „–" = befejezetlen tagmondat.



5. ábra. Összesítő grafikon a tekintetmintázatok eredményeiről

5.4. A tekintetviselkedés együttállásai

A leírt tekintetmintázatokat a kommunikáció multimodális jellege miatt érdemes együttjárásaik szerint is megvizsgálni. Vessük össze az ismétlések alatti és az ismétlések utáni tekintetirányokat. A rendelkezésre álló adatok alapján az ismétlés alatti és az ezt követő első tartalmas szó tekintetviselkedésének hat lehetséges kategóriája különíthető el:

a) TM + TP = ismétlés alatt a tekintet a partner irányától eltérő irányba mutat, míg a soron következő első tartalmas szó elhangzásakor a beszélő a partnerre tekint;

b) TP + TM = a beszélő az ismétlés alatt a partner irányába tekint, de a tartalmas szó alatt más irányba;

c) TP + TP = a beszélő végig az interjúvezető irányába tekint;

d) TM + TM = a beszélő sem az ismétlés elhangzásakor, sem az azt követő tartalmas szó alatt nem néz a partnerre;

e) TM + Ø = a beszélő tekintete nem a partnerre szegeződik az ismétlés alatt, viszont nem követi őt ugyanazon tagmondatban tartalmas szó, amelynek tekintetmintázatát elemezhetnénk;

f) TP + Ø = a beszélő a partnerre tekintve megismétli a szót, de az ismétlést tartalmas szó nem követi.

A leggyakrabban előforduló együttállás a TM + TP (az összes vizsgált jelenség 47%-a (45 eset)), vagyis, amikor a beszélő másfelé tekint az ismétlés alatt, de az első lényeges tartalmas szó kiejtésekor, mintegy megerősítésként, a partnerre tekint. Ez utalhat arra, hogy a szóban forgó ismétlések valamilyen információban gazdag vagy fontos szót előznek meg.

Az a) pontban található együttállás ellenkezőjére (TP + TM), vagyis amikor a beszélő az ismétlés elhangzása után tekintetét elkapja a hallgatóról, nem találtunk példát a vizsgált hat beszélőnél, ami két dolgot jelenthet. Egyrészt azt, hogy az ismétlés nem valamilyen hiba korrigálása, azaz nem egy kognitív önellenőrzési folyamat jelölője, hanem időhúzó tényező. Ebben az utóbbi esetben eredményünk alátámaszthatja Németh (2012) következtetését, amely szerint az ismétlés feladata az időhúzás. Másrészt várható az is, hogy ha az ismétlés során a beszélő nem bizonytalanodik el, tehát tekintete a partnerre szegeződik, akkor a soron következő ismétlés sem fog produkciós nehézséget eredményezni, amely miatt a beszélő hirtelen elkapná tekintetét az interjúvezetőről.

A TP + TP (27 eset, 28%), illetve TM + TM (24 eset, 25%) jelenségek, mint láthatjuk, közel azonos arányban fordultak elő. Az előbbi esetben a hallgató számára nem okoz problémát a beszéd követése és szemantikai tartalmának értelmezése, mivel a beszélő tekintetével figyelemmel követi a passzív fél reakcióit. Az utóbbi eset viszont kísértetiesen hasonlít a telefonon keresztül zajló spontán beszélgetésekre, amely szituációkban a hallgató nem látja az aktuálisan beszélő fél tekintetét, az ismétlés mégsem számít megakasztó tényezőnek.

Az utolsó csoportra (f) nem volt példa, tehát a tartalmas szó előhívása nélküli nem szándékolt ismétléskor a beszélő mind a 10 esetben más irányba nézett (TM + Ø).

5.5. Az ismétlések viszonya további multimodális markerekkel

A 96 tekintet-együttállást megvizsgáltuk a gesztusmintázatok szempontjából is. Arra voltunk kíváncsiak, hogy a személyközi kommunikációban milyen más modalitások játszanak még közre – a tekintet mellett – az elhangzott információadó tartalmas szavak hangsúlyos jellegének érzékeltetésében. Először a 24 esetet adó TM + TM tekintetmintázat-pár során észlelhető gesztusok vizsgálatainak eredményeit közöljük. Azokban az esetekben, amikor az interjúalany tekintete az ismétléstől kezdve az első tartalmas szóval bezárólag a partnertől eltérő irányba irányul, a kézi gesztusokat és a fejmozgást figyelve a következő kapcsolatokat detektáltuk: a kézi gesztusok és a fejmozgás (oldalra/előre irányuló fejbiccentés, bólogatás, fejrázás) önállóan, és együttesen is kísérhetik az ismétlést és az első tar-

talmas szót. Összesen 18 kézi gesztust és 9 fejmozgást azonosítottunk. Fejmozgást önállóan 2 esetben sikerült észlelnünk, kézi gesztikulációt pedig 11 esetben. Vagyis 7 esetben a két jelenség együttesen kísérte az ismétlést és az első tartalmas szót, és 5 esetben nem volt vizuálisan értékelhető mozgás (6. ábra).

6. ábra. TM+TM esetben megnyilvánuló gesztusmintázatok. Az elhangzott ismétlések: „de <de> amúgy tényleg a tanár az legalább"; „hogy <hogy> igen Petit el szeretném vinni egy sörre"; „viszont mi a <a> <a> lány legjobb barátnőjével csináltunk egy új baráti kört"; „meg <meg> <meg> van az emberben egy ilyen igény arra"

Feltehetjük ugyanakkor, hogy a gesztusok nem csak a fent elemzett esetben segítik a feldolgozást a hallgató részéről, hanem akkor is, amikor a beszélő tekintetének iránya a partner felé irányul (7. és 8. ábra). Ezek alátámasztásához azonban további mérési adatok és vizsgálatok szükségesek.

Az itt bemutatott eredményeket a nem szándékos ismétlések nem verbális jelei alapján kaptuk. Természetesen a multimodalitásnak része a beszédhangokkal való manipulálás is, beleértve a beszéddallam, az intenzitás és a tempó (benne a szünet) eszközeinek a használatát. Jól észlelhetően a prozódia ilyen jellegű használata is jelen van vizsgált anyagunkban, pontos adatokat azonban csak további mérések alapján tudunk közölni. Az ismétlések multimodális kifejezésének itt bemutatott eredményei ugyanakkor jól alátámasztják Hunyadi (2011a) azon állítását, hogy kézzelfogható kapcsolat van a verbális és a nem verbális jelenségek között.

7. ábra. Gesztusmintázatok, amikor a beszélő a tartalmas szó produkciója során a partnerre tekint. Az elhangzott tagmondatok a képek sorrendjében: „hogy <hogy> kitaláltam ezt"; „hát <hát> az egyszerű volt"; „de <de> látom az ilyen jeleket"; „azok a <azok> az amerikaiak"; „hát <hát> előkapták a pillangókést"

8. ábra. Nincs kézi gesztikuláció, csak oldalirányú fejbillentés („ez <ez> a balesetem volt")

Ez a rövid esettanulmány tehát arról tanúskodik, hogy a HuComTech korpusz hat interjúalanyának ismétlései és az ezt követő tartalmas szavak kiejtése alatti tekintetviselkedés és gesztikuláció egymással összefügg. Az ismétléseket követő tartalmas szavak memóriából történő előhívásához, illetve nyomatékosításához a beszélők olyan vizuális elemeket használnak, mint a tekintet irányával való manipulálás, kézi gesztikuláció és fejmozgások. Az esettanulmányban vizsgált verbális kifejezések produkciója és a hozzá tartozó mozdulatok közötti viszony azt

sugallja, hogy lehetséges összefüggés az ismétlések bizonyos típusait követő egyes gesztusok megléte és a nyelvi produkció mögötti kognitív működés között.

6. A HuComTech korpusz szintaktikai annotációs szintje

A következőkben a szintaktikai annotációs szint bemutatására, a spontán beszéd nyelvi megnyilatkozásainak mondatokra és tagmondatokra való felosztására kerül sor.

6.1. A szintaktikai annotációs szint célkitűzése és szabályrendszerének kialakítása

A szintaktikai annotációs szint létrehozásának kettős célja van: egyrészt létre kívánunk hozni egy olyan eszközt, melynek segítségével adott szempontok alapján rendszerszerűen le lehet írni a magyar beszélt nyelv nyelvtanát, másrészt – mivel ez az szint egyike a HuComTech multimodális korpusz annotálási szintjeinek – ezáltal hozzá kívánunk járulni ahhoz, hogy kutathatóvá váljék a beszélt nyelv szintaxisának a kommunikáció egyéb moduljaival való összefüggése is. A prozódiával való összefüggésének vizsgálata hozzájárulhat a beszéd teljesebb automatikus felismeréséhez, a prozódiai és a gesztusokra vonatkozó információkkal együtt pedig az ember–ember, valamint az ember–gép kommunikáció jobb megértéséhez juthatunk.

Az annotációs alapelvek kidolgozása során az empirikus adatok gyűjtése és osztályozása együtt járt az összegyűjtött és osztályozott adatok egyfajta preteoretikus rendszerben való összegzésével. Az annotációs szabályrendszer tehát induktív és deduktív módszer együttes alkalmazásával jött létre. Megközelítésünk preteoretikus volta egyrészt tükrözi a különböző leírások és elméletek közötti konszenzust, ugyanakkor nem szándékszik olyan elméleti elkötelezettséget tenni, ami az eredményül kapott annotáció későbbi széles körű felhasználását korlátozhatná.

6.2. A szintaktikai szabályrendszer elemzési alapegységei (az annotáció szintaxisának alapjai)

Elemzésünk lényegét tekintve strukturális és nem funkcionális. A szintaktikai jelölési szabályrendszer keretei között központi fogalomnak a tagmondatot tekintjük. A tagmondat alapvető kritériumaként a predikatív viszonyt tesszük fel. Az

ennél szélesebb értelmű mondatot önmagában nem, csak a tagmondatok kapcsolódásain keresztül határozzuk meg. Ennek értelmében egy mondat ott ér véget (és potenciálisan ott kezdődik egy újabb mondat), ahol további tagmondatot már nem tudunk strukturálisan csatlakoztatni. A mondat belső szerkezeti összefüggései között fontosnak tartjuk a hierarchia és a szerkezeti hiány azonosítását és jelölését. A szerkezeti hierarchia esetében a tagmondatok közötti alá-, fölé- és mellérendelést jelöljük. A hiány fogalmának bevezetését a spontán nyelvi megnyilatkozások természete kívánja meg. Az egyes kapcsolódásokban azon szerkezeti elemek hiányát vizsgáljuk, amelyek az adott szerkezet építését befolyásolják. A szintaktikai szerkezetek ilyen szempontú besorolása lehetővé teszi, hogy világosan látható legyen, adott esetekben a grammatikalitás milyen fokon teljesült explicit módon.

Nem célunk a mondatoknak funkcionális elemzését adni, különösen azért, mert a spontán beszédben történő nyelvi megnyilatkozás funkciója – nem kis mértékben éppen a hiányok miatt – gyakran áttevődik nyelven kívüli eszközökre vagy akár jelöletlen is marad, így a megnyilatkozás funkcióinak rendszerszerű megismerése hangsúlyozottan az interpretáció feladata lenne. Ezért – mivel ezt számos esetben a struktúra nem jelöli – a tagmondatok közötti hierarchikus viszonyok jelölésén túli részletesebb elemzésre, így pl. az alá- és mellérendelő mondattípusok megnevezésére nem vállalkozunk.

Ami a hiányzó mondatelemeket illeti, a következő megfontolásokkal élünk. Csak azoknak az elemeknek a felszíni hiányzását tüntetjük fel szerkezeti hiányként is, amelyek vonzatok, azaz elhagyhatatlan bővítmények (vö. Komlósy 1992; Keszler 2000), tehát a grammatikai struktúra sérülése nélkül nem hagyhatók el azon nyelvi egység mellől, amelyhez tartoznak. Az alanyt viszont nem tekintjük vonzatnak, elmaradása mégis hiányként van jelölve annak ellenére is, hogy a személyragból következtethetünk rá.

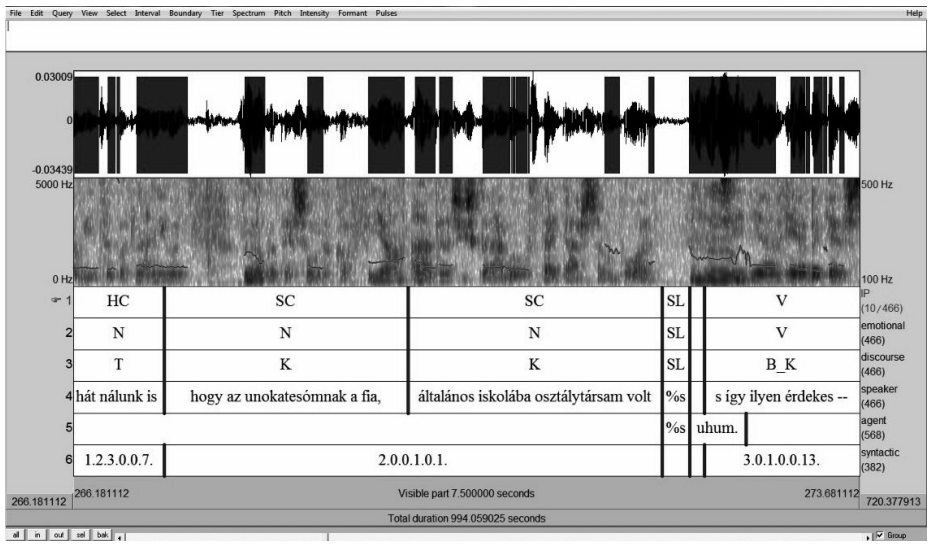
6.3. Az annotációs szabályrendszer módszere és a kódrendszer

Módszerünk egyrészt a kategorizálás, másrészt a formalizálás. Ábrázoljuk az egyes szintaktikai egységek sorrendiségét, mégpedig úgy, hogy – mivel szigorúan az elhangzott beszédben dokumentált tényekből indulunk ki – szabályrendszerünk megengedi az írott nyelvre vonatkozó leíró grammatika szempontjából nem megfelelő sorrend létrejöttét, létezését. Így megengedi a leíró nyelvtani szempontból „helyes" és „helytelen" mondatokat, és nem jelöli azok ilyen szempontú megkülönböztetését. Alapvető célja, hogy azonosítsa a strukturálisan meghatározható szintaktikai viszonyokat a mondatszerkezetek felépítésében (valójában a szavak egymáshoz való kapcsolódásaiban) és a tagmondatok egymásutániságában.

A kódrendszer a szintagmatikus összefüggéseket ábrázolja, a mondatrészi kategóriák egyszerű függőségi viszonyait, de megmutatja a lehetőséget arra is, hogy hogyan reprezentálhatjuk magát a kontextust. Az annotációs szabályrendszer hiánykategóriájának használatával csak a leíró nyelvtan szintaxisa szempontjából értelmezünk valóságos hiányokat és így „kivételt képező", netalán „helytelen" mondatszerkezeteket. Ezzel szemben a multimodális szinteken történő szimultán annotálás valójában éppen azt teszi lehetővé, hogy megfigyelhessünk, észrevegyünk és azonosítsunk egy adott modalitásból hiányzó elemet mint annak egy másik modalitásban és ugyanazon időpillanatában történő megvalósulását: így például egy szintaktikai szinten befejezetlen mondat lezárását nyomon követhetjük a hozzá kapcsolódó kézmozdulatok, mimika és egyéb jelek figyelembe vételével.

A szintaktikai kapcsolódások kódolásához számozást alkalmazunk. A több számjegyű számozás a tagmondatok közötti sorrendiséget, a tagmondatok közötti viszonyt és a fellépő szintaktikai hiány jellegét fejezi ki. Az első számjegy a mondat kezdetét jelöli, és az ezt követő számjegyek az egymást követő tagmondatok egymás közötti, valamint saját belső szintaktikai viszonyaira – beleértve a hiányokat is – utalnak. Ez a számozás ott fejeződik be, ahol további szerkezeti kapcsolatokat nem találunk. Így ez a pont valójában egybeesik azzal, amit egy hagyományos értelemben vett mondat végének neveznénk (és ami az ezt követő nyelvi anyagot egy újabb mondat kezdeteként jelöl meg). Így számunkra egy mondat határait nem az interpretáció és nem a prozódia határozza meg, hanem a – bármilyen hiányos – szintaktikai viszonyok.

A 9. ábrán láthatjuk a kódolási rendszert, a multimodális annotáció hatodik, legalsó szintjén. A kódrendszerben az első számjegy tehát az adott mondaton belül az egymást követő tagmondatok sorszámát jelenti. A második számjegy azt jelöli, hogy az adott tagmondathoz tartozik-e alárendelés, és ha igen, akkor ez a hányadik számú tagmondat. Ha ilyen alárendelés nincs, akkor ez 0 értékkel van jelölve. A harmadik számjegy a tagmondathoz tartozó mellérendelő tagmondat(ok) sorszámát jelöli. Ha ilyen nincs, akkor ott a 0 érték szerepel. A negyedik számjegy azt mutatja meg, hogy az adott tagmondat hányadik számú tagmondatnak az alárendeltje. Ennek hiányában itt is a 0 érték jelenik meg. (Az alárendelt kapcsolatban álló tagmondatok kölcsönös meghatározottsága lehetővé teszi, hogy e viszony jelölését függetlenné tegyük a tagmondatok felszíni sorrendjétől.) Az ötödik számjegy a grammatikai kapcsolat hiányát jelöli, rámutat a beágyazás és a beékelés jelenségeire. A hatodik számjegy a hiány kategóriáit hozza felszínre (főmondat, előtte vagy utána álló mellérendelő tagmondat, utalószó, kötőszó,



9. ábra. A szintaktikai annotációs szint kódolása

grammatikai, illetve logikai alany, állítmány, tárgy [tágyas ige esetén], határozó, jelző, ige), illetve jelöljük azt is, ha nem hiányzik semmi az adott tagmondatból, vagy ha a mondat befejezetlen, illetve ha irreleváns a hiány kategóriájának felvetése.

Jelölési konvencióinknak megfelelően az egymást követő, különböző elemzési szempontokra vonatkozó számok, jegyek közé pontot teszünk. Ha egy elemzési szemponthoz több számérték is tartozik, akkor az azokra utaló számjegyeket vesszővel választjuk el.

Tudatában vagyunk annak, hogy az itt ismertetett annotációs rendszer nem ad választ a legrészletesebb viszonyokra irányuló szintaktikai kérdésekre, és különösen nem elégíti ki a funkcionális nyelvleírás mentén felvetődő számos igényt. Mindez a választott megközelítésünk következménye. A szándékunk csupán annyi volt, hogy olyan preteoretikus annotálást adjunk, amelyből sokféle, egymástól különböző elméleti megközelítés is haszonnal kiindulhat, és főként azt várjuk, hogy a spontán beszéd szintaxisára is kiterjesztett annotált multimodális adatbázisunkkal a nyelvtechnológia, különösen a szintaxist valamilyen formában figyelembe venni szándékozó beszédfelismerés és -szintézis számára nyújthatunk közvetlenül is felhasználható anyagot.

7. A HuComTech korpusz pragmatikai szempontból: unimodális annotáció

A HuComTech korpusz szintaktikai annotációjának ismertetése után annak következő szintje, a pragmatikai szempontú annotáció kerül bemutatásra. A 7. pont az unimodális, funkcionális megközelítésű pragmatikai annotáció, a 8. pont pedig a multimodális pragmatikai annotáció mögött meghúzódó elméleti megfontolásokat, valamint az annotáció szintjeit és címkéit mutatja be. A HuComTech-projekt arra is lehetőséget nyújt, hogy a különösen a fizikai jelek felismerését és szintézisét előtérbe helyező video- és audioannotáció mellett vállalkozzunk egy olyan annotációra is, ami egy kommunikatív esemény pragmatikai viszonyait tárja fel. Ennek során ugyancsak figyelembe veszünk audio- és videojeleket, azonban ezekre az előzőktől eltérően tekintünk. Nem magukat a jeleket keressük és – ha megtaláltuk – annotáljuk, hanem az esemény pragmatikai vonatkozásait szem előtt tartva pragmatikai jellemzőket keresünk, és ezekhez rendeljük magukat a jeleket. Azaz ez az annotálás az előzővel ellentétben kifejezetten interpretatív, így a kétféle annotálás szükségszerűen feltételezi és kiegészíti egymást. Ez történhet unimodálisan éppúgy, mint multimodálisan. Mivel a kétféle megközelítés különböző elméleti elvárásokon alapszik, korpuszunkon mindkétféle megközelítést megvalósítjuk.

Az itt következőkben az unimodális funkcionális–pragmatikai annotációra térünk ki, és annak eszközét, alszintjeit és címkéit mutatjuk be. Ezt megalapozandó, röviden ismertetjük a séma kidolgozása mögött húzódó elméleti megfontolásokat és gyakorlati célkitűzéseket.

7.1. Elméleti megfontolások

Pragmatikai szempontból éppúgy, mint a technológia követelményeinek szem előtt tartása miatt céljaink közé tartozik a kommunikatív esemény bizonyos, unimodálisan is jól megragadható mozzanatainak kinyerése a HuComTech korpuszból, majd pedig a felismert jegyek, markerek alapján a társalgás menetével kapcsolatos predikciók megtétele. A technológiai szempont itt különösen jelentős: míg egy hétköznapi kommunikatív esemény során az lehet a benyomásunk, hogy az esemény mozzanatait holisztikusan, azaz az adott pillatanban elérhető összes (verbális és nem verbális) modalitás együttes feldolgozásával érzékeljük és értelmezzük, a technológiai implementáció megkívánja, hogy ezen összetett adatfolyamot jól kezelhető diszkrét elemeire bontsuk. Maga a vállalkozás azonban pragmatikaelméleti szempontból is fontos lehet, hiszen a kommunikáció mul-

timodálisan igencsak összetett eseményét így felbontva az értelmezés mögöttes kognitív folyamataira is fény derülhet.[9]

Több olyan alkalmazott nyelvészeti, társalgáselemzési megalapozottságú tanulmány született már az interperszonális kommunikáció invariáns struktúrájának és szekvenciális elrendezésének, rendezőelveinek megragadásával kapcsolatban, amely a kézenfekvő verbalitásra helyezve a hangsúlyt, multimodálisan értelmezte a kommunikációt (Sacks et al. 1974; Sacks 1995; Németh T. 1996; Schegloff 2006; Abuczki 2011), ugyanakkor igen kevés olyan kutatás folyt és tanulmány született, amely különválasztva a modalitásokat, unimodális megközelítést alkalmazva, csupán egyetlen modalitásból érkező információkra szorítkozna (Esposito–Esposito 2010). Kutatócsoportunk úgy véli, hogy a társalgás komputációs megragadásának és a dialógusrendszerek modellezésének érdekében érdemes modalitásonként külön-külön explicitté tenni az egyes kommunikatív jelenségek gépi eszközökkel is detektálható felszíni jegyeit.

Az általunk a multimodális annotáció mellett, attól különböző eljárásként alkalmazott unimodális pragmatikai annotációtól azt várjuk, hogy lehetőséget biztosítson arra, hogy az ember–ember és az ember–gép kommunikációt egyaránt új megvilágításba helyezzük. Ennek megfelelően a jelenleg folyó annotálás során új megközelítést alkalmazva különválasztjuk az információs csatornákat (percepciósan bármilyen nehéznek is tűnik mindez). Célunk a kommunikációs esemény bizonyos aspektusainak megragadása pusztán vizuális, vagy pusztán akusztikus input alapján. Az unimodális annotáció mögött meghúzódó feltevésünk az, hogy a kommunikációs eseménynek vannak olyan összetevői, amelyek pusztán vizuális eszközökkel is megragadhatóak, és egyetlen modalitás markerei is kifejezhetnek bizonyos kommunikatív funkciókat, ráadásul akár olyanokat is, amelyek bizonyos mértékben különböznek azoktól, amelyeket ugyanezen modalitás más modalitásokkal való kombinációjában kifejez.

Az annotáció során bár markereket annotálunk, közvetlenül nem egy-egy tekintetmintát vagy fejmozdulatot stb. keresünk, hanem kommunikatív funkciókat, és ezekhez azonosítjuk az őket megvalósító markereket (Hunyadi 2011a). Azaz ezen funkciókat valamilyen markerek segítségével azonosítjuk, de a kiindulás mégis a vélt kommunikációs esemény megragadása, ellentétben az eddigi videoannotációk fő vonulatával, ahol lényegében előre meghatározott fizikai markerek (pl. arckifejezések, kézmozdulatok, testtartás) jelenlétét annotáltuk, függet-

[9] Ahogy McNeill fogalmaz: A gesztusok „a beszélt nyelvvel időben, jelentésükben és funkciójukban oly szorosan összefonódnak, hogy a beszélt nyelvi megnyilatkozásokat és a gesztusokat akár ugyanazon mögöttes mentális folyamat különböző oldalainak is tekinthetnénk" (McNeill 1995, 1).

lenül azok funkciójától. Nem gondoljuk azt, hogy unimodálisan teljes biztonsággal meg tudjuk határozni ezeket a funkciókat, ezeket a multimodális verifikálás felül is bírálhatja. Megközelítésünk a kommunikáció formális alapszerkezetének megismerésére irányul, úgy, hogy ezáltal egy tetszőleges kommunikációs esemény formális modellen alapuló technológiai létrehozása lehetővé váljék (vö. Hunyadi 2012). Ahhoz, hogy a nem verbális modalitások formális leírása egységes modellben legyen kezelhető, előbb külön-külön kell az egyes modalitásokat megvizsgálni. A Hunyadi-modell (2011b) lehetővé teszi a verbális és a nem verbális kommunikáció modalitásonként különválasztott absztrakt és felszíni jegyeinek együttes technológiai reprezentálását, ezzel elősegítve a több kommunikációs csatorna által közvetített információ egyidejű feldolgozását. Unimodális annotációnk tehát a fenti gondolatmenetet követve a kommunikáció absztrakt és felszíni jegyeinek rendszerbe foglalásához, ezen belül különösen a beszélőváltás és az egyetértés/nem-egyetértés gépileg is detektálható jegyeinek szisztematikus leírásához kíván hozzájárulni.

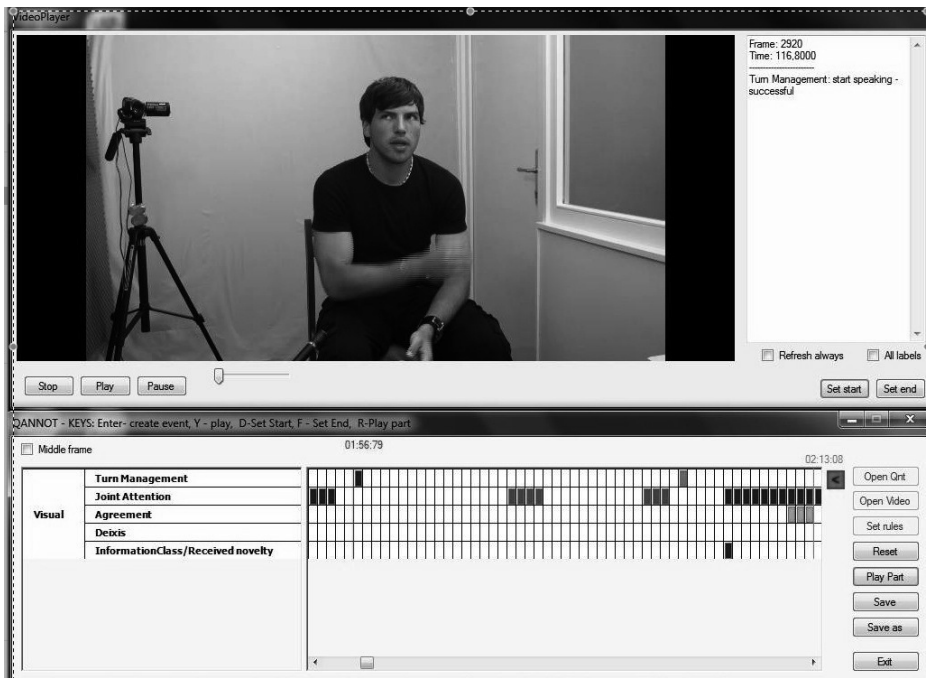
7.2. Az annotáció folyamata

Az annotáció folyamata gyakorlatilag a szemlélt kommunikációs esemény időbeli határokkal *(timestamps)* való ellátása. Ezen határokon belül egy legördülő menüből kiválaszthatjuk a mozdulathoz vagy a gesztusként értelmezhető mozdulatsorhoz illeszkedő címkét, és jelöljük az adott esemény időbeli terjedelmét (l. 10. ábra). A különböző típusú eseményekhez rendelt színkódolt sávok egymással időben szinkronizáltak, így szinkrón természetű (nem szekvenciális), ún. együttállásra vonatkozó címkelekérdezéseket is lehetővé tesznek.

Unimodális annotációs rendszerünk előnye, hogy olyan nyelv- és kultúrafüggetlen, univerzális kategóriákkal dolgozik, mint például a beszélés kezdete és vége által határolt társalgási forduló, vagy az unimodálisan is jól megragadható egyetértés/nem egyetértés gyakori kommunikatív aktusa. Mindezek a kategóriák a beszélők korától, társadalmi státuszától és a szituációtól függetlenül is invariáns részét képezik mindenféle társalgásnak, ami lehetővé teszi unimodális sémánk bármely nyelvű spontán beszéden vagy multimodális korpuszon való alkalmazását.

Az unimodális annotáció történhet vagy csak vizuális, vagy csak audió input alapján is. Itt az előbbiről szólunk részletesebben.

Vizuális-unimodális annotációnk első áttekintése megerősíti, hogy pusztán vizuális input (gesztikuláció, szájmozgás) alapján is meg tudjuk állapítani, hogy ki az aktuális beszélő fél az interakcióban. A megnyilatkozások időtartamának



10. ábra. Az unimodális annotáció felhasználói felülete a Qannot programban

megoszlása alapján azt is meg tudjuk állapítani, hogy mennyire kiegyensúlyozott az interakció a résztvevő felek között. Továbbá, csupán egyetlen modalitás alapján a másik félre fordított figyelem és a beszélgetésbe vonódás mértékét is meg tudjuk állapítani. Sőt, a nem verbális viselkedés vizuális jegyeiből kitűnik az is, amikor a beszélőnek csak az a szándéka, hogy megkezdje a mondandóját, de nem tudja elindítani azt, mert félbeszakítják, vagy az, amikor valamilyen közelebbről nehezen azonosítható okból nem folytatja, esetleg újrakezdi a beszélést.

A fentieknek megfelelően az unimodális annotáció a következő szinteken történik (az angol elnevezések az annotáció valóságos címkéinek felelnek meg):[10]

1. a fordulókezelés szintje *(Turn Management Class),* amelyen belül megkülönböztetjük a beszélés kezdetét a beszélés végét, a közbevágást és a beszédkezdés szándékát;
2. a figyelem szintje *(Attention Class),* amelyen belül megkülönböztetjük a közös figyelembe vonódást és a figyelemfelkeltést;
3. az egyetértés szintje *(Agreement Class),* amely kétféle attribútumot, pozitívat vagy negatívat vehet fel, és azon belül is megkülönbözteti a teljes egyetértést, a részleges egyetértést,

[10] A HuComTech-projekt kutatásának egésze angol nyelven dokumentált, így annak annotációs sémája is nemzetközi, angol nyelvű terminológiát követ.

az egyetértés alapértelmezett esetét, a bizonytalanságot; a nem-egyetértés alapértelmezett esetét, a társalgás elvágásának/blokkolásának szándékát és az érdektelenséget/közönyt;

4. a deixis szintje *(Deixis Class),* amelyben olyan deiktikus viselkedés kerül rögzítésre, amelyre a videoannotáció megfelelő szintje nem tér ki;
5. az információ szintje *(Information Class),* amelyben új információ észlelését rögzítjük.

A HuComTech-kutatócsoport által alkalmazott angol nyelvű unimodális annotációs terminológia rövid összegzése a következő (l. 1. táblázat):

1. táblázat. Az unimodális annotációi szintjeinek és címkéinek áttekintő táblázata

Unimodális annotáció osztályai *(classes)*		**Unimodális annotáció címkéi *(attribútumai)***
Turn Management Class (fordulókezelés/beszélőváltás)		start speaking successfully (sikeres beszédkezdés, beszélőváltás)
		breaking in (közbevágás)
		intending to start speaking (beszédkezdés szándékának kifejezése, de nem sikeres beszédkezdés
Attention Class (figyelem)		calling attention (figyelemfelhívás)
		paying attention (figyelem kifejezése)
Agreement Class (egyetértés)	+agreement (egyetértés)	default case of agreement (egyetértés alapértelmezett esete)
		full agreement (teljes egyetértés)
		partial agreement (részleges egyetértés)
		uncertainty (bizonytalanság)
	–agreement (nem egyetértés)	default case of disagreement (nem egyetértés alapértelmezett esete)
		blocking (társalgás elvágása)
		uninterested (közönyös)
Deixis Class (deixis)		other (egyéb, a videoannotáció szintjén nem jelölt deixis)
Information Class (információ-struktúra)		received novelty (észlelt új információ)

Nemcsak kommunikatív, hanem komputációs szempontból (Bunt–Black 2000) is érdemes a fordulót (*turn;* l. unimodális sémánk első szintjét) a kommuniká-

ció alapegységének tekinteni, hiszen vizuálisan és akusztikusan is jól körülhatárolható, többségében univerzális jelenségek kísérik a forduló átadását, így a beszélőváltás számítógépes eszközökkel is jól detektálható. A fordulókezelés a társalgásban a küldő szerep jogának a beszélők közötti elosztását jelenti. A fordulókezelés és a beszélőváltás gépileg is detektálható jegyeinek (nyelvfüggetlen és nyelvfüggő verbális, vizuális és nem verbális akusztikus jegyeinek) explicit felfedése és szisztematikus rendszerbe (pl. döntési fába) foglalása elsődleges céljaink közé tartozik. Egyik kérdésünk az, hogy a beszélgetőpartnerek hogyan osztják ki maguk között a szó átvételének jogát, illetve hogyan, milyen vizuális és akusztikus jelek (markerek) alapján ismerik fel a beszélőváltásra alkalmas pillanat elérkezését. Ez a kérdés akkor is eldönthető, ha a Qannot videokép alatti hangerőszabályozóján elnémítjuk a felvételt (vagyis eldönthető unimodálisan is), ugyanis látható, mikor kezdi el a beszélő a beszélést és mikor hagyja abba (ráadásul kitűnik, hogy a beszélés kezdetének vizuális markere gyakran hamarabb jelenik meg az akusztikus markernél). Annotációink lekérdezésével többek között azt is meg tudjuk állapítani, hogy átlagosan milyen hosszúságú észlelt szünet jelenti a forduló átadását, és ezzel milyen nem verbális jelenségek járnak együtt. Unimodális annotációnk eredményeinek segítségével további kihívásaink közé tartozik olyan jól körülhatárolható és gépi eszközökkel is megragadható kommunikatív viselkedés automatikus felismerése, mint például az egyetértés és a nem egyetértés kommunikatív aktusa. Eredményeinkkel remélhetőleg hozzájárulhatunk a beszédtechnológia és a dialógusrendszer-modellezés egyik feladatának, a társalgási fordulóvég, más szóval a lehetséges beszélőváltási pont predikciójának megoldásához is, amely predikció alapját kell hogy képezze mindenféle természetes menetű, gördülékeny, időben szinkronizált kérdés–válasz, vagy bármely egyéb kommunikációs szekvenciát követő ember–gép kommunikációnak.

7.3. Várható eredmények

Az unimodális annotáció befejezésével és kiértékelésével a következő elméleti és empirikus eredményeket várjuk elérni:

1. a kommunikációs esemény szerkezetének pontosabb feltárása;
2. a fordulókezelés és beszélőváltás jellemzőinek pontosabb, modalitásonkénti megragadása;
3. az interakciót irányító, diskurzust szervező verbális akusztikus jelenségek, a nem verbális akusztikus és a vizuális viselkedés további részleteinek és azok összefüggéseinek a megismerése;
4. hozzájárulás egy beszélőváltást előrejelző program betanításához.

A következő pontban a különféle információs csatornákból érkező információkat együttesen figyelembe vevő pragmatikai annotáció, a multimodális pragmatikai annotáció bemutatására kerül sor.

8. A HuComTech korpusz pragmatikai szempontból: multimodális annotáció

8.1. A multimodális pragmatikai annotáció célja

A HuComTech korpusz multimodális pragmatikai annotációjának célja kettős. Az elmélethez kötődő cél felfedni a hétköznapi személyközi kommunikáció mögöttes, hierarchikus és szekvenciális szerkezeti sajátosságait, amelyek a kommunikációs események strukturálásában alapvető szerepet játszanak. Ezt a célt úgy tudjuk megvalósítani, hogy a kommunikatív viselkedésekben rejlő verbális akusztikus, nem verbális akusztikus, valamint vizuális jegyeket, Hunyadi (2011a; 2012) terminológiájában markereket, kommunikatív funkciójuk szerint azonosítjuk, valamint az azonosított markereket korreláltatjuk egymással (illetve a többi annotációs szint releváns címkéivel).

8.2. Az annotációs rendszer

A multimodális pragmatikai annotáció alapját a kommunikatív aktusok képezik. A kommunikatív aktusok multimodális illokúciós aktusokként értelmezendők, mivel a verbális közlesek mellett a gesztusokat és a nem verbális akusztikus információkat is figyelembe vesszük az annotáció során. A társalgás szerkezetében a fordulók a legkarakteresebb egységek (és a fordulók kommunikatív aktusokból állnak), valamint a kommunikatív aktusok képesek a beszélő hétköznapi vágyainak és szándékainak kifejezésére is, ezért kézenfekvő volt az annotációs rendszer kommunikatív aktusokra történő alapozása. Tipológiánk kidolgozásához a Bach- és Harnish-féle rendszert választottuk ki (Bach 2006). E megkülönböztetés melletti érveinket részletesen tárgyaljuk Abuczki (2011)-ben. A kommunikatív aktusok típusai:

- konstatívok *(constatives)* = ítélkezők: válaszadás, megerősítés, informálás, predikció, visszaemlékezés
- direktívák *(directives)* = végrehajtók: kérés, parancs, javaslattétel
- komisszívok *(commissives)* = elkötelezők: beleegyezés (pl. egy fogadásba), fölajánlás, ígéret
- viselkedők *(acknowledgements)*: üdvözlés, búcsúzás, elfogadás (pl. meghívásé)
- nem azonosítható *(none)*

A négy kommunikatívaktus-típus közül a konstatívok olyan aktusokat foglalnak magukban, amelyek a beszélőnek egy propozicionális tartalomhoz fűződő hiedelmét fejezik ki. Mégpedig úgy, hogy a beszélő mindeközben szándékozza azt is, hogy az aktus propozicionális tartalmát feldolgozza és elhiggye a hallgató is (Abuczki 2011). A direktívek olyan aktusokat tartalmaznak, amelyeknek 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 tegye meg a cselekedetet (uo.). A komisszívok olyan aktusok, amelyek a beszélő azon szándékát fejezik ki, amellyel elkötelezi magát egy jövőbeli aktus megtételére (uo.). Végezetül a viselkedők olyan aktusok, amelyek a beszélő valamilyen affektív, érzelmi, attitűdbeli viszonyulását fejezik ki a hallgató iránt (uo.). Léteznek azonban olyan esetek is, amikor egy fordulóban nem azonosíthatók a kommunikatív aktusok imént tárgyalt típusai. Ez az eset akkor áll fenn, amikor az adott forduló nem tartalmaz olyan információt, amely fogódzóként szolgálhatna a fent ismertetett négy típusba való besorolásához. Ezeket a részeket a „none" (nem azonosítható) címkével jelöljük az annotáció során.

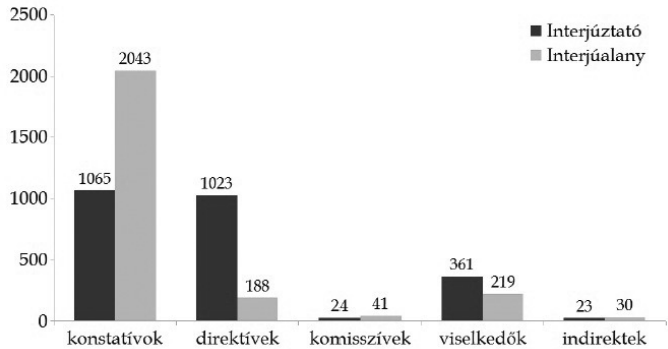
A kommunikatív aktusok mellett az úgynevezett támogató aktusokat is annotáljuk. Ezek az aktusok kiegészítik, támogatják a velük egységben szereplő kommunikatív aktust. A tematikus kontroll szintjén azt vizsgáljuk, hogy a társalgás egyes fordulói milyen módon illeszkednek a társalgás egészébe. A fordulók efféle globálisabb vizsgálata megmutatja, hogy a társalgás során az egyes témák miképpen szerveződnek egységekbe, hogyan történik az egyes társalgási témák motivált egymásba fűzése, illetve a motiválatlanság. A társalgási témák motivált egymásba fűzése rávilágít a társalgásbeli együttműködés mintázataira is.

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 jár együtt.

8.3. Előzetes eredmények

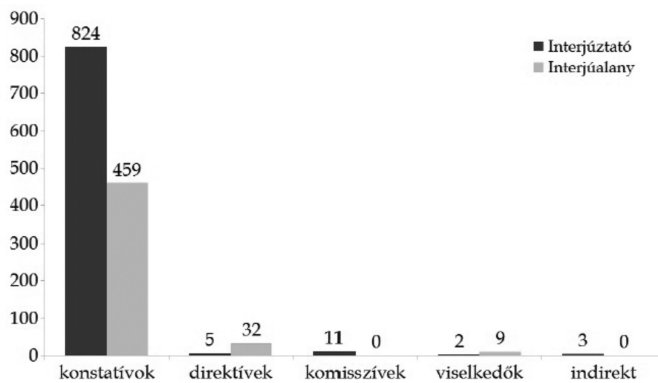
Noha a HuComTech korpusz multimodális pragmatikai annotációja jelenleg is zajlik, előzetes eredményeink jól mutatják, hogy a kommunikatív aktusok, a társalgási fordulók, valamint a fordulókból álló szomszédsági párok közötti viszonyok rendszerszerűek. 35 formális és informális felvétel annotációja alapján láthatjuk, hogy az interjúvezető és az interjúalany szerepeiknek megfelelő kommunikatívaktus-típusokat hoztak létre (11. ábra): az interjúvezető a szcenáriónak megfelelően kb. fele-fele arányban produkál direktív és konstatív aktusokat, míg

az interjúalanyok elsősorban válaszolnak a direktív aktusokra, így az ő konstatív aktusaik száma kiemelkedően magas.



11. ábra. A különböző kommunikatívaktus-típusok előfordulásainak száma

Előzetes címkelekérdezéseink azt is mutatják, hogy a támogató aktusok közül a visszajelzések *(backchannelek)* alapvetően a konstatív aktusokkal járnak együtt (12. ábra).

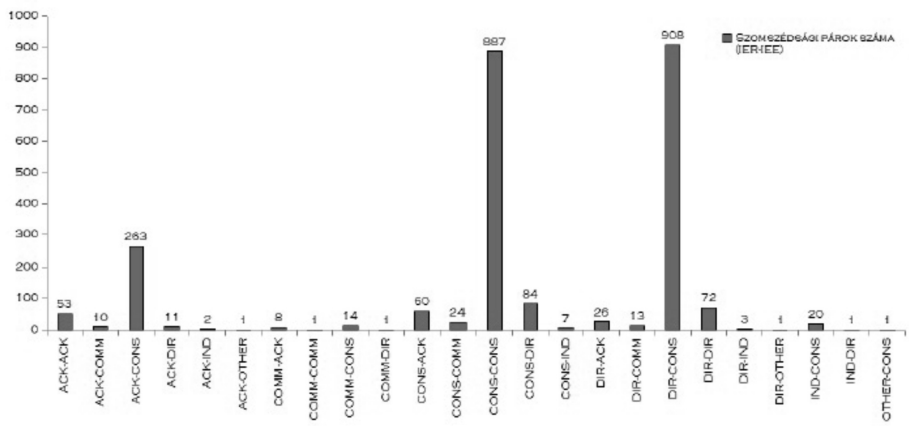


12. ábra. A különböző kommunikatívaktus-típusokra adott visszajelzések száma

Ennek oka minden bizonnyal az, hogy a konstatív aktusok esetében megjelenik egy propozicionális tartalom, amelynek megértéséről a hallgató biztosítja a beszélőt. Fontos azonban látni, hogy a visszajelzések funkciói ennél sokkal szerteágazóbbak, így további kutatások alapját kell képezniük. A 12. ábrán azt is láthatjuk, hogy alapvetően az interjút készítő személy a gyakoribb visszajelző fél a társalgásban. Ennek oka egyrészről az lehet, hogy a felvételek forgatókönyvében

az interjúalany gyakrabban kerül a társalgás középpontjába (aktív kommunikátori szerepben), mint az interjúkészítő fél.

Megvizsgáltuk az egyes szomszédsági párokat alkotó fordulók kommunikatív aktusait is (13. ábra). Az egyes fordulók párokba rendeződésének két kiemelkedő mintázatát figyelhetjük meg: a konstatív–konstatív, valamint a direktív–konstatív együttállást. Ez a kétféle mintázat alapvetően strukturálja a korpusz diskurzusszerkezetét, s jól tükrözi a társalgás mögött meghúzódó forgatókönyvet: az interjúkészítő feladata az információ kérése, az interjúalany feladata az információ adása (ez történik a direktív–konstatív együttállás során). A konstatív–konstatív együttállás valószínűsíthetően egy-egy társalgási téma kidolgozásában játszik fontos szerepet, ám a korpusz jelenlegi feldolgozottsága mellett ilyen típusú lekérdezést még nem tudunk végrehajtani.



13. ábra. A kommunikatív aktusok szomszédsági párba rendeződése

8.4. Kitekintés

A HuComTech korpusz multimodális pragmatikai annotációs rendszerének több előnye is van más annotációs rendszerekhez képest:

1. A rendszer nyelvfüggetlen, univerzális kategóriákkal dolgozik. Mind a kommunikatív aktusok típusai, mind a támogató aktusok, mind a tematikus kontroll tulajdonságai univerzális jellemzői a társalgásnak.

2. Az annotáció során együttesen, multimodálisan vesszük figyelembe a vizuális, a nem verbális akusztikus, valamint a verbális akusztikus információkat. Mivel annotációnk kevés lexikai információn alapul, lehetővé válhat a későbbiekben az információk automatikus kinyerése.

3. A címkézés lehetővé teszi a későbbiekben, hogy egy-egy típust önmagában hívjunk le, és szükség eseten tovább finomítsunk. Ez a megoldás gazdaságos.

4. A fordulók mint strukturális elemek és a kommunikatív aktusok típusai mint funkcionális elemek együttes szerepeltetése lehetővé teszi, hogy a fordulókból kibontakozó szomszédsági párokhoz is tudjunk megfelelő kommunikatívaktus-típusokat rendelni. Ez közelebb vihet bennünket olyan predikciók megtételéhez, amelyeknek a segítségével anticipálni tudjuk a következő fordulót a társalgásban.

Ennek a rendszernek a hatékonyságát jól jelzik előzetes eredményeink, amelyek empirikus adatokat szolgáltatnak a társalgásbeli szerepek jellemzőinek fölfejtéséhez, így hozzájárulva a forgatókönyvvel kapcsolatos predikciók megtételéhez.

9. Összefoglalás

A HuComTech korpusz létrehozását eredetileg egy elméleti cél elérése motiválta: az, hogy létrehozzuk az ember–gép kommunikáció olyan technológiai modelljét, amely alapvetően épít az ember–ember kommunikáció lényeges és e feladat szempontjából releváns jellemzőire. A feladat ilyen megközelítését a bevezetőben indokoltuk.

A 2. pontban javaslatot tettünk az ember–gép kommunikáció újszerű elméleti–technológiai modelljére. A szemléletében generatív, főbb jellemzőiben moduláris, szekvenciális és multimodális modell olyan primitívek halmazát tételezi fel, amelyek, vagy amelyeknek műveletekkel létrehozott képződményei valamilyen módon felszíni alakzatot is öltenek ún. markerek formájában. A modell lényeges feladata ugyanakkor, hogy olyan keretet biztosítson, amely nemcsak a kommunikáció elméleti leírását adja, hanem lehetővé teszi a technológiai megvalósítást is. Ehhez szükség van arra, hogy a markerek tulajdonságait a technológia számára elérhető módon leképezzük.

Az annotáció feladata, hogy általa a elméleti modell építőköveit, a markereket feltárjuk. Így az annotáció során markereket (vagy bizonyos szintjein a markerek előfordulásának funkcionális–pragmatikai interpretációit) keresünk és azonosítunk. Az annotáció azon túlmenően, hogy alátámasztja és elősegíti a kommunikáció felépítésének a megismerését, alapjául szolgál a technológiai megvalósításnak is. Ezt a modell technológiai interfésze biztosítja azáltal, hogy

a modell elméleti komponensének markerekben megjelenő kimenetét a modell technológiai komponensében paraméterek értékeinek felelteti meg. A modell fontos jellemzője, hogy kétirányú, azaz egyaránt szolgálja a szintézist (egy kommunikatív esemény technológiai megvalósítását) és az analízist (ezen esemény interpretációját, „megértését"). Lehetővé teszi e két, ellentétes irányú folyamat egyidejű kezelését, ami által alkalmassá válik egy ember–gép kommunikáció kétirányú folyamatának egységes kezelésére.

A tanulmány további pontjaiban az annotálás különböző lényeges aspektusait mutattuk be, ismertetve az annotálás folyamatát és az adatbázis lekérdezése alapján már elérhető bizonyos eredményeit.

A 3. pontban ismertettük a videoannotálás elveit és az annotálás során megjelölt főbb kommunikációs jegyeket. Bemutattuk, milyen mértékben befolyásolhatja az egyes jegyek (markerek) multimodális együttállása a kommunikatív esemény interpretációját.

A 4. pontban az akusztikus információ kódolásáról szóltunk. Kitértünk a manuális kódolás egyes eredményeire, és az általa felvetett, a további feldolgozásra hatással levő kérdéseit is indokoltuk, valamint részleteztük az automatikus prozódiai annotálás munkálatait. Az így kinyert prozódiai jellemzők birtokában lehetővé válik a vizsgált kommunikációs események gépi detektálásának vagy akár predikciójának az algoritmizálása, ami fontos lépés lehet nyelvtechnológiai alkalmazások továbbfejlesztésében.

A 5. pont esettanulmánya azt mutatta be, hogyan alkalmazható a multimodális annotáció a kommunikációs események jobb megértésére. Nem szándékos ismétlések és az erre következő tartalmas szavak kiejtése során megfigyeltük a csatlakozó tekintetviselkedést és gesztikulációt. Bemutattuk, hogy a beszélők az ismétléseket követő tartalmas szavak memóriából történő előhíváshoz, illetve nyomatékosításához jellemzően olyan vizuális elemeket használnak, mint a tekintet irányával való manipulálás, kézi gesztikuláció és fejmozgások. Az esettanulmány során láthatóvá vált egyes gesztusok és verbális kifejezések produkciója közötti összefüggés. Ezek az ismeretek ugyancsak jól hasznosíthatóak lehetnek a kommunikáció bizonyos mozzanatainak gépi felismerésében és értelmezésében.

A 6. pont a beszélt nyelv szintaxisának rendszerszerű lejegyzését célul kitűző újszerű szintaktikai annotálás elveit mutatta be számos, az írott nyelv szintaxisa által nehezen kezelhető eset bemutatásával. Ettől az annotálástól azt várjuk, hogy egyrészt jobban megértsük a beszélt nyelv szintaktikai szerveződését, másrészt hozzájáruljunk a szintaxisra támaszkodó nyelvtechnológiai alkalmazások továbbfejlesztéséhez.

A 7. és a 8. pontban a pragmatikai annotálás két párhuzamos, egymástól független rendszerét mutattuk be. Az unimodális annotálás (7. pont) célja, hogy

pragmatikailag értelmezhető kommunikatív tulajdonságokat ismerjünk fel csupán egyetlen, mégpedig a vizuális csatorna jelei alapján. Első eredményeink azt mutatják, hogy – ellentétben a természetesnek tűnő elvárással – az ilyen unimodális megközelítés során is lehetővé válik bizonyos kommunikatív funkciók felismerése, ami különösen fontos lehet egy-egy esemény modalitásonkénti technológiai megvalósítása számára. A multimodális pragmatikai annotáció (8. pont) is jelentős újdonságot mutat. Előnyként értékelhetjük más annotációs rendszerekhez képest, hogy nyelvfüggetlen, univerzális kategóriákkal dolgozik, és kevés lexikai információn alapul, amelyek a későbbiekben tovább finomíthatók. A kommunikatív aktusok típusai és a funkcionális elemek együttes szerepeltetésével lehetővé válik a társalgás következő fordulójának az anticipálása.

Összegzésképpen tehát megállapíthatjuk, hogy az annotált korpusz alapján létrehozott – és rövidesen közvetlenül is elérhető – adatbázis célja a kommunikáció rendszerszerű megismerésén túlmenően olyan eszköz nyújtása a nyelvtechnológusok és más szakemberek számára, amely lehetővé teszi a multimodális emberi viselkedés jobb és rendszerszerű megértését, valamint olyan alkalmazások létrehozását, amelyek az ember–gép kommunikáció szerteágazó területein – a humán felhasználónak az eddigieknél érezhetően emberközelibb környezetet biztosítva – saját hatékonyságát jelentősen növelheti.

Irodalom

Abuczki Ágnes 2011. A multimodális interakció szekvenciális elemzése. In: Németh T. (2011, 119–144).

Alessandro, Christophe d' – Piet Mertens 1995. Automatic pitch contour stylization using a model of tonal perception. Computer Speech and Language 9: 257–288.

Bach, Kent 2006. Speech acts and pragmatics. In: Michael Devitt – Richard Hanley (szerk.): Blackwell guide to philosophy of language. Malden MA & Oxford: Blackwell. 147–156.

Bernsen, Niels Ole – Laila Dybkjær 2005. Natural and multimodal interactivity engineering – Directions and needs. In: Kuppevelt et al. (2005, 1–22).

Bódog Alexa (szerk.) 2011. Az ember–gép kommunikáció technológiájának elméleti alapjai. IKUT zárókötet. Debrecen: Debreceni Egyetemi Kiadó.

Boersma, Paul – David Weenink 2005. Praat: Doing phonetics by computer. (Version 5.1.43)

Bunt, Harry – William Black 2000. The ABC of computational pragmatics. In: Harry Bunt – William Black (szerk.): Abduction, belief and context in dialogue: Studies in computational pragmatics. Amsterdam & Philadelphia: John Benjamins. 1–46.

Chomsky, Noam 1965. Aspects of the theory of syntax. Cambridge MA: MIT Press.

Chomsky, Noam 1977. Essays on form and interpretation. New York: North Holland.

Chomsky, Noam 1986. Knowledge of language: Its nature, origin and use. New York: Praeger.

Cole, Ronald A. – Joseph Mariani – Hans Uszkoreit – Annie Zaenen – Victor Zue (szerk.) 1997. Survey of the state of the art in human language technology. Cambridge: Cambridge University Press.

Dix, Alan J. – Janet E. Finlay – Gregory D. Abowd – Russell Beale 2003. Human–computer interaction. 3rd edition. Englewood Cliffs, NJ: Prentice Hall.

Esposito, Anna – Antonietta Maria Esposito 2010. On speech and gestures synchrony. In: Anna Esposito – Alessandro Vinciarelli – Klára Vicsi – Catherine Pelachaud – Anton Nijholt (szerk.): Lecture notes in computer science. Berlin: Springer. 252–272.

Flanagan, James L. 1997. Overview – Multimodality. In: Cole et al. (1997, 329–342).

Fox, Barbara – Makoto Hayashi – Robert Jasperson 1996. Resources and repair: A cross-linguistic study of syntax and repair. In: Elinor Ochs – Emanuel A. Schegloff – Sandra A. Thompson (szerk.): Interaction and grammar. Cambridge: Cambridge University Press. 185–237.

Földesi András 2011. Unimodális funkcionális annotáció a HuComTech multimodális korpuszban. In: Bódog (2011, 40–46).

Gósy Mária 2002. A megakadásjelenségek eredete a spontán beszéd tervezési folyamatában. Magyar Nyelvőr 126: 192–204.

Gósy Mária 2008. A zaj hatása a beszédre. Beszédkutatás 2008: 5–21.

Grice, H. Paul 1957. Meaning. Philosophical Review 67: 377–388.

Grice, H. Paul 1975. Logic and conversation. In: Peter Cole – Jerry L. Morgan (szerk.): Syntax and semantics, vol. 3: Speech acts. New York: Academic Press. 41–58.

Gyarmathy Dorottya 2011. A multimodális interakció szekvenciális elemzése. In: Németh T. (2011, 119–144).

Hunyadi László 2011a. A multimodális ember–ép kommunikáció technológiái – elméleti modellezés és alkalmazás a beszédfeldolgozásban. In: Németh T. (2011b, 15–42).

Hunyadi László 2011b. Az ember–gép kommunikáció elméleti-technológiai modellje. Háttér és alapkérdések. In: Bódog (2011, 6–12).

Hunyadi, László 2012. Multimodal human–computer interaction technologies – Theoretical modeling and application in speech processing. Argumentum 7: 240–260.

Jakobson, Roman 1969. Nyelvészet és poétika. In: Roman Jakobson (szerk.): Hang – jel – vers. Budapest: Gondolat Kiadó. 211–258.

Jong, Nivja H. de – Tor Wempe 2009. Praat script to detect syllable nuclei and measure speech rate automatically. Behavior Research Methods 41: 385–390.

Kenesei István 2000. Szavak, szófajok, toldalékok. In: Kiefer Ferenc (szerk.): Strukturális magyar nyelvtan 3. Morfológia. Budapest: Akadémiai Kiadó. 75–136.

Keszler Borbála (szerk.) 2000. Magyar grammatika. Budapest: Nemzeti Tankönyvkiadó.

Komlósy András 1992. Régensek és vonzatok. In: Kiefer Ferenc (szerk.): Strukturális magyar nyelvtan 1. Mondattan. Budapest: Akadémiai Kiadó. 299–527.

Kuppevelt, Jan C. J. van – Laila Dybkjær – Niels Ole Bernsen (szerk.) 2005. Advances in natural multimodal dialogue systems (Text, Speech and Language Technology 30). Dordrecht: Springer.

Lewis, David K. 1969. Convention. Cambridge MA: MIT Press.

Mariani, Joseph 1997. Multimodality. In: Cole et al. (1997, 329–370).

McNeill, David 1995. Hand and mind: What gestures reveal about thought. Chicago: University of Chicago Press.

Mertens, Piet 2004. The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model. In: Bernard Bel – Isabelle Marlien (szerk.): Proceedings of the 2nd International Conferemce of Speech Prosody, Nara, 23–26 March 2004.

Németh Zsuzsanna 2012. A javítási műveletek interakciós funkciói: ismétlés és csere a magyarban. Beszédkutatás 2012: 154–167.

Németh T. Enikő 1996. A szóbeli diskurzusok megnyilatkozáspéldányokra tagolása (Nyelvtudományi Értekezések 142). Budapest: Akadémiai Kiadó.

Németh T. Enikő 2011a. A humán kommunikáció modelljei és az ember–gép kommunikáció. In: Németh T. (2011b, 43–62).

Németh T. Enikő (szerk.) 2011b. Ember–gép kapcsolat. A multimodális ember–gép kommunikáció modellezésének alapjai. Budapest: Tinta Könyvkiadó.

Oviatt, Sharon L. 2003. Multimodal interfaces. In: Julie A. Jacko – Andrew Sears (szerk.): The human–computer interaction handbook: Fundamentals, evolving technologies and emerging applications. Mahwah, NJ: Lawrence Erlbaum. 286–304.

Pápay Kinga – Szeghalmy Szilvia – Szekrényes István 2012. HuComTech multimodal database annotation. Argumentum 7: 330–347.

Sacks, Harvey 1995. Lectures on conversation. Cambridge MA & Oxford: Blackwell.

Sacks, Harvey – Emanuel A. Schegloff – Gail Jefferson 1974. A simplest systematics for the organization of turn-taking for conversation. Language 50: 696–735.

Schegloff, Emanuel A. 2006. Sequence organization in interaction: A primer in conversation analysis. Cambridge: Cambridge University Press.

Shannon, Claude – Warren Weaver 1949. The mathematical theory of communication. Urbana, IL: University of Illionis Press.

Sperber, Dan – Deirdre Wilson 1986/1995. Relevance: Communication and cognition. Cambridge, MA & Oxford: Blackwell.

Szekrényes István – Csipkés László – Oravecz Csaba 2011. A HuComTech-korpusz és -adatbázis számítógépes feldolgozási lehetőségei. Automatikus prozódiai annotáció. In: Tanács Attila – Vincze Veronika (szerk.): A VIII. Magyar Számítógépes Nyelvészeti Konferencia előadásai. Szeged: Szegedi Tudományegyetem. 190–198.

Thórisson, Kristinn R. 2008. Modeling multimodal communication as a complex system. In: Ipke Wachsmuth – Manuela Lenzen – Günther Knoblich (szerk.): Springer lecture series in computer science: Modeling communication with robots and virtual humans. New York: Springer. 143–168.

Tóth Csilla 2011. Tekintetmintázatok és funkcióik a HuComTech-projekt szimulált állásinterjúiban. In: Németh T. (2011b, 101–118).

Wahlster, Wolfgang 1991. User and discourse models for multimodal communication. In: Joseph W. Sullivan – Sherman W. Tyler (szerk.): Intelligent user interfaces. New York: ACM Press. 45–67.

Wahlster, Wolfgang (szerk.) 2006. SmartKom: Foundations of multimodal dialogue systems. New York: Springer.

A theoretical–technological model of human–computer interaction and its implications for language technology

Abstract: With its goal to contribute to a more humanlike human-machine interaction, the HuComTech project aims at studying and describing those aspects of human-human communication that are supposed to be of primary relevance for our interaction with machines. Whereas language is undoubtedly the most important prerequisite of communication, it is by far not the only one: gestures, gaze patterns, head and body movements as well as ways of speech production equally contribute to a successful communication being essentially multimodal. The paper gives an overview of this interdisciplinary approach to communication presenting a novel two-way, generative model of communication at the intersection of theory and technology and offering the first observations and results based on 60 hours of audio-video recordings. In particular, it describes principles of the video annotation (and observations on facial expressions and their alignment with other multimodal features), principles of the audio annotation (and observations on speech tempo, intonation, repetitions), principles of the annotation of the syntax of speech as well as of the uni- and multimodal pragmatic annotation. The paper also includes a case study of observed gaze patterns and their communicative functions.

Keywords: human–computer interaction, multimodality, multimodal corpus, annonation, prosody, syntax, pragmatics